

Phase transition in inhomogenous Erdős-Rényi random graphs via tree counting

Ghurumuruhan Ganesan *

New York University, Abu Dhabi

Abstract

Consider the complete graph K_n on n vertices where each edge e is independently open with probability $p_n(e)$ or closed otherwise. Here $\frac{C-\alpha_n}{n} \leq p_n(e) \leq \frac{C+\alpha_n}{n}$ where $C > 0$ is a constant not depending on n or e and $0 \leq \alpha_n \rightarrow 0$ as $n \rightarrow \infty$. The resulting random graph G is inhomogenous and we use a tree counting argument to establish phase transition in G . We also obtain that the critical value for phase transition is one in the following sense. For $C < 1$, all components of G are small (i.e. contain at most $M \log n$ vertices) with high probability, i.e., with probability converging to one as $n \rightarrow \infty$. For $C > 1$, with high probability, there is at least one giant component (containing at least ϵn vertices for some $\epsilon > 0$) and every component is either small or giant. For $C > 8$, with positive probability, the giant component is unique and every other component is small. As a consequence of our method, we directly obtain the fraction of vertices present in the giant component in the form of an infinite series.

Key words: Erdős-Rényi random graphs, phase transition, tree counting argument.

AMS 2000 Subject Classification: Primary: 60J10, 60K35; Secondary: 60C05, 62E10, 90B15, 91D30.

*E-Mail: gganesan82@gmail.com

1 Introduction

Consider the complete graph K_n on n vertices $\{1, 2, \dots, n\}$ and corresponding edges e_1, \dots, e_m where $m = \binom{n}{2}$. An edge e_i is independently open with probability $p_n(e_i)$, and closed otherwise. Throughout we assume that

$$\frac{C - \alpha_n}{n} \leq \min_{1 \leq i \leq m} p_n(e_i) \leq \max_{1 \leq i \leq m} p_n(e_i) \leq \frac{C + \alpha_n}{n} \quad (1.1)$$

where $0 \leq \alpha_n \rightarrow 0$ as $n \rightarrow \infty$. The resulting random graph G is an Erdős-Rényi (ER) graph (Janson et al (2000)). Strictly speaking G is one version of an ER random graph and in the original paper, Erdős and Rényi (1960) have studied phase transition in a slightly different model of random graphs.

When $p_n(e_i) = \frac{C}{n}$ for all i , the probabilities for an edge begin open are the same and the resulting random graph is homogenous. Phase transitions in homogenous graphs have been studied in great detail and the common approaches include a combination of vertex exploration, random walks and branching processes (see Janson et al (2000), Durrett (2007), Alon and Spencer (2008) and references therein) and usually look at stochastic domination of the exploration process from above and below by branching processes. For a more recent comprehensive account, we refer to the monographs of Hofstad (2016). For example, the results related to phase transitions are treated in Chapter 4, Theorems 4.4-4.8 in Hofstad (2016).

When the edge probabilities are not the same, the resulting random graph G is inhomogenous and the above analysis is not directly applicable. We illustrate with an example in Section 2. To study phase transitions in inhomogenous random graphs, we use in this paper a combinatorial tree counting argument to estimate the sizes of small and mid size components. To make the proof self contained, we give the proofs for both subcritical and super critical regimes.

Tree counting arguments have been used before in the analysis of random graphs. For example, Bollobas (2001) (Chapter 7), Durrett (2007) (Chapter 2) and Alon and Spencer (2008) (Chapter 11) have all used tree counting arguments in various forms for different purposes. Bollobas (2001) uses the tree counting with coarser estimates to analyse the connectivity regime of homogenous random graphs where each edge is independently open with probability $p = \frac{C \log n}{n}$ for some constant $C > 0$. Durrett (2007) obtains estimates on the number of tree components of a fixed size using the tree

counting argument for $p = \frac{C}{n}$. Alon and Spencer (2008) use the tree counting argument with $p = \frac{C}{n}$ for comparison with a Poisson branching process.

Our argument is different from the above in the sense that we obtain more precise tree counting estimates for the small component sizes when edges are open with probability of the order of $\frac{1}{n}$ (see Lemmas 3.1 and 3.2, Section 3). We use the above estimates to indirectly deduce the presence of a giant component. An important consequence of our method is that for the supercritical regime of $C > 1$, we obtain the analytical solution for the fraction $q(C)$ of vertices present in giant components in the form of an infinite series. For details, we refer to the proof of Theorem 4, Section 8.

Model Description

We briefly describe the probability space first. For $1 \leq i \leq m$, we define the state of the edge $e_i \in K_n$ on the probability space $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ where $\Omega_i = \{0, 1\}$, $\mathcal{F}_i = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$ and \mathbb{P}_i denotes the Bernoulli probability measure with $\mathbb{P}_i(\{1\}) = p_n(e_i) = 1 - \mathbb{P}_i(\{0\})$. Thus open edges are assigned a value of 1 and closed edges are assigned a value of 0. We define the random graph G on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\Omega = \otimes_{i=1}^m \Omega_i$, $\mathcal{F} = \otimes_{i=1}^m \mathcal{F}_i$ and $\mathbb{P} = \otimes_{i=1}^m \mathbb{P}_i$.

Throughout the paper, we study open components of the graph G and we give a brief description. Let e_i and e_j , $j \neq i$ be two edges in K_n . We say e_i and e_j are *adjacent* if they share an endvertex. We say that a sequence of distinct edges $P = (f_1, f_2, \dots, f_k)$, $f_i \in \{e_j\}$ is a *path* in K_n if the edge f_i is adjacent to the edge f_{i+1} for every $1 \leq i \leq k-1$. We say that P is an *open path* (in G) if P is a path and every f_i , $1 \leq i \leq k$, is open. Let v_1 be the endvertex of the edge f_1 not common to f_2 and let v_k be the endvertex of f_k not common to f_{k-1} . We say that v_1 and v_k are *endvertices* of the path P .

Let $1 \leq i \neq j \leq n$ be fixed. We say that vertex i is connected to vertex j by an open path if there is an open path P_{ij} containing i and j as endvertices. Let \mathcal{E}_i be the set of all vertices v , $1 \leq v \leq n$, connected to the vertex i by an open path. We assume that $i \in \mathcal{E}_i$ and define \mathcal{E}_i to be the *open component* of the graph G containing the vertex i . We also refer to open components simply as components.

For $C > 0$, let

$$\delta = \delta(C) := C - 1 - \log C. \quad (1.2)$$

We have that $\delta(1) = 0$ and $\delta(C) > 0$ for $C \neq 1$. We are interested in the size of components of G for the case $C < 1$ (subcritical) and the case $C > 1$

(supercritical) which we discuss separately below.

Subcritical case

We have the following result for $C < 1$. For $1 \leq i \leq n$, let \mathcal{E}_i denote the open component of the graph G containing the vertex i defined in the previous subsection. For integer $M \geq 1$ let

$$H_1 = H_1(M) := \bigcap_{1 \leq j \leq n} \{\#\mathcal{E}_j \leq M \log n\} \quad (1.3)$$

denote the event that the size of every component is at most $M \log n$. Throughout, the size of a component refers to the number of vertices present in the component.

Theorem 1. *Fix $C < 1$ and let $\delta = \delta(C) > 0$ be as in (1.2). Fix $M > \frac{1}{\delta}$ and let $\theta < \min(M\delta - 1, 1)$ be fixed. There is a positive constant $L_1 = L_1(C, M, \theta)$ so that*

$$\mathbb{P}(H_1(M)) \geq 1 - \frac{1}{n^\theta} \quad (1.4)$$

for all $n \geq L_1$.

Thus with high probability, (i.e., with probability converging to one as $n \rightarrow \infty$), every component has size at most a constant multiple of $\log n$.

For values of C slightly farther away from the critical value of one, we have stronger rates of decay.

Theorem 2. *Fix $C < e^{-1}$ and $\theta > 0$. There are positive constants $M = M(C, \theta)$ and $L_2 = L_2(C, \theta, M)$ so that*

$$\mathbb{P}(H_1(M)) \geq 1 - \frac{1}{n^\theta} \quad (1.5)$$

for all $n \geq L_2$.

Supercritical case

To study the supercritical case of $C > 1$, we first have some preliminary definitions. For integer $r \geq 1$, let T_r denote the number of labelled trees on r vertices. We set $T_1 = 1$ and for $r \geq 2$, we recall the Cayley formula (see van Lint and Wilson (2010)) for T_r as $T_r = r^{r-2}$ for $r \geq 2$. In this paper, we

do not require the use of the exact Cayley formula and therefore, we simply treat T_r as a sequence.

For $C > 0$, define

$$q = q(C) := \sum_{r=1}^{\infty} \frac{T_r e^{-r}}{C(r-1)!} e^{-\delta r} = \sum_{r=1}^{\infty} \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!} \quad (1.6)$$

where $\delta = \delta(C) \geq 0$ is as in (1.2). For the supercritical case, the function $q(C)$ plays a crucial role in determining the fraction of vertices in the giant component of G .

The following result collects together the important properties of $q(C)$ needed for the proofs.

Theorem 3. *The function $q(C)$ satisfies the following properties:*

(i) *For $C > 0$, we have*

$$0 < q(C) \leq 1. \quad (1.7)$$

(ii) *If $C \leq 1$, then*

$$q(C) = 1. \quad (1.8)$$

(iii) *For $C > 1$, the function $q(C)$ is strictly decreasing in C and*

$$0 < q(C) < 1. \quad (1.9)$$

(iv) *For $C > 0$, we have that $q = q(C)$ satisfies the relation*

$$q = e^{-C(1-q)}. \quad (1.10)$$

Using a combinatorial approach, like for e.g. generating functions, we obtain the property (iv) that $q = q(C)$ satisfies the relation (1.10) (see e.g. van Lint and Wilson (2001), Grimmett (1980)). The term q satisfying (1.10) is also the extinction probability of a Poisson branching process with mean offspring size C (see Durrett (2007)). Thus the expansion $q(C)$ in (1.6) is the analytical solution for the extinction probability. We give a probabilistic proof of properties (i) – (iii) in Section 7 using the properties of random graphs.

Fix $C > 1$ and $\epsilon > 0$. For $i \geq 1$, let

$$Y_i := \mathbf{1}(\#\mathcal{E}_i > \epsilon n) \quad (1.11)$$

denote the indicator function of the event that vertex i belongs to a giant component, i.e., a component containing more than ϵn vertices. We then have

$$Y_n(\epsilon) := \sum_{i=1}^n Y_i \quad (1.12)$$

denotes the sum of sizes of all giant components. We use the term giant component to roughly mean any component which contains at least a constant fraction $f \in (0, 1)$ of the n vertices. We make the notion precise based on the context.

For $0 < \gamma < 1$ and define the event

$$V(\gamma, \epsilon) := \{(1 - q(C) - \gamma)n \leq Y_n(\epsilon) \leq (1 - q(C) + \gamma)n\}. \quad (1.13)$$

For $M > 0$ let

$$H_2 = H_2(M, \gamma, \epsilon) := V(\gamma, \epsilon) \bigcap_{1 \leq j \leq n} \left(\{\#\mathcal{E}_j \leq M \log n\} \bigcup \{\#\mathcal{E}_j > \epsilon n\} \right) \quad (1.14)$$

denote the event that $V(\gamma, \epsilon)$ occurs and every component is either giant or small. Here we say that a component is small if its size is at most $M \log n$. We have the following result.

Theorem 4. *Fix $C > 1$ and let $\delta = \delta(C) > 0$ be as in (1.2). Fix $M > \frac{1}{\delta}$ and $0 < \gamma < 1$. There exists $\epsilon_1 = \epsilon_1(C, M, \gamma) > 0$ such that the following holds for all $0 < \epsilon < \epsilon_1(C)$. There is a positive constant $L_3 = L_3(C, M, \gamma, \epsilon)$ so that*

$$\mathbb{P}(H_2) \geq 1 - \gamma \quad (1.15)$$

for all $n \geq L_3$.

Since $q(C) < 1$ for $C > 1$ and $0 < \gamma < 1$ is arbitrary, we have with arbitrarily large probability that there exists a giant component of G .

Uniqueness of the giant component

In Theorem 4, we have established that for any $C > 1$, with high probability there is at least one giant component, i.e., a component whose size is larger than ϵn for some constant $\epsilon > 0$. We now see that for values of C slightly farther away from one, the giant component is unique with positive probability.

For $0 < \gamma < 1$ and for $1 \leq i \leq n$, let

$$V_i = V_i(\gamma) := \{(1 - q(C) - \gamma)n \leq \#\mathcal{E}_i \leq (1 - q(C) + \gamma)n\} \quad (1.16)$$

denote the event that the size of the component \mathcal{E}_i containing the vertex i is in the range $[(1 - q(C) - \gamma)n, (1 - q(C) + \gamma)n]$. Here $q(C)$ is as in (1.6). For $M > 0$ define

$$H_3 = H_3(M, \gamma) := \bigcup_{i=1}^n V_i \bigcap \bigcap_{1 \leq j \leq n, j \neq i} \{\#\mathcal{E}_j \leq M \log n\} \quad (1.17)$$

to be event that there exists a unique giant component with size in the range $[(1 - q(C) - \gamma)n, (1 - q(C) + \gamma)n]$ and the size of every other component is at most $M \log n$. For positive numbers C, ϵ and ω , let $\delta_0(C, \epsilon, \omega) = \delta(C) - C\epsilon - \omega$. We also use the definition of δ_0 in future results.

Theorem 5. *There is a unique $C_0 > 1$ such that $\delta_0(C_0, \frac{1}{2}, 0) = 0$. We have that $C_0 < 8$.*

Fix $C > C_0$ and let $\delta = \delta(C) > 0$ be as in (1.2). Fix $M > \frac{1}{\delta}$ and let $\alpha > 0$ be such that $\delta_0(C, \alpha + \frac{1}{2}, 0) > 0$. Fix $0 < \gamma < 1$. There is a positive constant $L_4 = L_4(C, \alpha, \gamma, M)$ so that

$$\mathbb{P}(H_3(M, \gamma)) \geq 1 - e^{-\alpha C} - \gamma \quad (1.18)$$

for all $n \geq L_4$.

For any $C > 8$, we therefore have with positive probability that there exists a unique giant component.

The paper is organized as follows: In Section 2, we provide an example of an inhomogenous random graph. The three following sections obtain preliminary estimates. In Section 3, we collect the tree counting estimates for non giant components, i.e., components with size at most ϵn for some constant $\epsilon > 0$. In Section 4, we obtain estimates on the mid size components whose size lies in the range $[M \log n + 1, \epsilon n]$ for some constant $M > 0$. We then obtain mean and variances estimates for $Z_n(\epsilon) = n - Y_n(\epsilon)$ (see (1.12)) in Section 5 needed for the proofs of the main Theorems.

Using the preliminary estimates of the above Sections, we first prove Theorem 1 regarding the subcritical case $C < 1$ in Section 6. We then provide a probabilistic proof for properties (i) – (iii) of Theorem 3 in Section 7. In Section 8, we prove the supercritical case $C > 1$ of Theorem 4 and finally, in Section 9, we prove Theorems 5 and 2.

2 Inhomogenous random graph example

Suppose that the edge probabilities $p_n(e) \in \{\frac{C-\alpha_n}{n}, \frac{C}{n}\}$ for all edges $e \in K_n$ and suppose that

$$\# \left\{ e \in K_n : p_n(e) = \frac{C - \alpha_n}{n} \right\} = \frac{1}{8}n^2.$$

We recall that K_n is the complete graph on n vertices. Thus there are $\frac{1}{8}n^2$ edges in K_n with $p_n(e) = \frac{C-\alpha_n}{n}$ and the rest have $p_n(e) = \frac{C}{n}$.

To compare homogenous and inhomogenous random graphs, we perform a coupling procedure as follows. Let $\{X(e)\}_{e \in K_n}$ be a set of independent and identically distributed (i.i.d) random variables uniformly distributed in $[0, 1]$ and defined on the probability space $(\Xi, \mathcal{G}, \mathbb{P}_c)$. For a particular realization of $\{X(e)\}_{e \in K_n}$, define the graphs

$$G^- = \left\{ e \in K_n : X(e) \leq \frac{C - \alpha_n}{n} \right\}$$

and

$$G = \{e \in K_n : X(e) \leq p_n(e)\}.$$

Thus G is the inhomogenous random graph where edge e is open with probability $p_n(e)$ independently of all the other edges. The graph G^- is the homogenous random graph obtained when every edge is independently open with probability $\frac{C-\alpha_n}{n}$.

The phase transition for the homogenous graph G^- essentially involves the study of two properties (Durrett (2007)). (a) The existence of a unique giant component containing roughly $(1 - q(C))n$ vertices and (b) No middle ground in the sense that every component other than the giant component has at most $D \log n$ vertices for some constant $D > 0$. As described in Section 1, the term $q(C)$ is the probability of extinction for the Poisson branching process.

By construction, the graph $G^- \subseteq G$. However, we argue below that even if the homogenous graph G^- satisfies the phase transition properties (a) and (b) above, the inhomogenous random graph G need not. We use the following two estimates.

(a1) Let

$$E_{dif} = \left\{ e \in K_n : \frac{C - \alpha_n}{n} < X_n(e) \leq p_n(e) \right\}$$

be the set of edges which are open in G but closed in G^- . Let R_{dif} be the number of edges in E_{dif} . If $\alpha_n = \frac{16}{\sqrt{n}}$, then we have that

$$\mathbb{P}_c(R_{dif} \in [\sqrt{n}, 3\sqrt{n}]) \geq 1 - \frac{2}{\sqrt{n}}$$

for all $n \geq 2$.

(a2) For $1 \leq i \leq n$, let d_i denote the degree of vertex i in the graph G . We have that

$$\mathbb{P}_c\left(\sup_{1 \leq i \leq n} d_i \leq 3 \log n\right) \geq 1 - \frac{D}{n^2}$$

for some constant $D > 0$ and for all $n \geq 2$.

From (a1) – (a2) we have that with high probability (i.e. with probability converging to one as $n \rightarrow \infty$), the random variable R_{dif} lies in the range $[\sqrt{n}, 3\sqrt{n}]$ and the maximum degree of a vertex in the graph G is at most $3 \log n$.

Proof of (a1) – (a2): For (a1) we proceed as follows. The random variable $\#E_{dif}$ is binomially distributed with parameters $m = \frac{1}{8}n^2$ and $p_b = \frac{\alpha_n}{n}$. For $\alpha_n = \frac{16}{\sqrt{n}}$, we have that the mean

$$\mathbb{E}_c(R_{dif}) = mp_b = \frac{\alpha_n}{n} \frac{1}{8} n^2 = 2\sqrt{n}$$

and the variance $var(R_{dif}) = mp_b(1 - p_b) \leq mp_b = 2\sqrt{n}$. Here \mathbb{E}_c denotes the expectation operator corresponding to the probability measure \mathbb{P}_c . Using the Chebychev inequality, we therefore have that

$$\mathbb{P}_c(|R_{dif} - \mathbb{E}R_{dif}| > \sqrt{n}) \leq \frac{var(R_{dif})}{n} \leq \frac{2\sqrt{n}}{n} = \frac{2}{\sqrt{n}} \quad (2.1)$$

This proves (a1).

We prove (a2) as follows. Suppose $d_i = \sum_{j=1}^n Z_{i,j}$ denotes the degree of vertex i , where $Z_{i,j} = 1$ if vertices i and j are joined by an edge and zero otherwise. The random variables $\{Z_{i,j}\}_{1 \leq j \leq n}$ are independent and for $1 \leq j \leq n$, we have

$$\mathbb{E}_c Z_{i,j} = p_{i,j} \leq p_u = \frac{C + \alpha_n}{n} \leq \frac{C + 1}{n}$$

for all $n \geq n_0$ large, using (1.1). The number n_0 does not depend on the choice of i or j . For $s > 0$ we therefore have

$$\mathbb{E}_c e^{sZ_{i,j}} = 1 - p_{i,j} + e^s p_{i,j} = 1 + (e^s - 1)p_{i,j} \leq 1 + (e^s - 1)p_u$$

and

$$\mathbb{E}_c e^{sd_i} = \prod_{j=1}^n \mathbb{E}_c e^{sZ_{i,j}} \leq (1 + (e^s - 1)p_u)^n \leq \exp((e^s - 1)(C + 1))$$

where the last estimate is obtained using $1 + x \leq e^x$ for $x > 0$.

Setting $s = 1$ and using the Markov inequality we have that

$$\mathbb{P}_c(d_i \geq 3 \log n) \leq \frac{\mathbb{E} e^{d_i}}{n^3} \leq \frac{D}{n^3}$$

for all $n \geq n_0$ and for $D = \exp((e - 1)(C + 1))$. Thus

$$\mathbb{P}_c \left(\bigcup_{1 \leq i \leq n} \{d_i \geq 3 \log n\} \right) \leq \frac{D}{n^2} \quad (2.2)$$

for all $n \geq n_0$ and so with high probability, every vertex in the graph G has degree at most $3 \log n$. This proves (a2). \blacksquare

We use properties (a1) – (a2) to compare the inhomogenous random graph G with the homogenous random graph G^- . Suppose that the term $R_{dif} \in [\sqrt{n}, 3\sqrt{n}]$ and every vertex in the graph G has degree at most $3 \log n$. By construction, we have that $G^- \subset G$ and the random variable R_{dif} is the number of edges which are open in G but closed in G^- . Even if the homogenous graph G^- satisfies the phase transition properties (a) and (b) above, the graph G could have the following component structure. The giant component in $G \supset G^-$ is the same giant component as in G^- . There is a component C_{mid} consisting of $x \in [\sqrt{n}, 3\sqrt{n}]$ edges all of which belong to \mathcal{E}_{dif} , i.e., the set of edges open in G but closed in G^- and the rest of the components of G have size at most $D \log n$.

Let $\#C_{mid}$ be the size, i.e., the number of vertices in the component C_{mid} . We have that

$$\frac{2\sqrt{n}}{3 \log n} \leq \#C_{mid} \leq 6\sqrt{n} \quad (2.3)$$

and so that graph G does not satisfy the phase transition properties (a) – (b). *Proof of (2.3):* For the lower bound, we use the fact that $\sum_{v \in C_{mid}} d_v = 2x \geq 2\sqrt{n}$ where d_v denotes the degree of vertex $v \in C_{mid}$. Since the degree of every vertex in G is at most $3 \log n$, we also have that $\sum_{v \in C_{mid}} d_v \leq 3 \log n \#C_{mid}$. Thus $\#C_{mid} \geq \frac{2\sqrt{n}}{3 \log n}$.

For the upper bound, we use the fact that there are $x \leq 3\sqrt{n}$ edges in C_{mid} and so the number of vertices in C_{mid} is at most $6\sqrt{n}$. ■

An analogous argument as above holds if we use the homogenous random graph $G^+ \supset G$ obtained by allowing every edge to be independently open with probability $\frac{C}{n}$.

3 Tree counting estimates

For $C > 0$ let $\delta = \delta(C)$ be as in (1.2). For $\omega > 0, \epsilon \in (0, 1)$ and $i = 0, 1, 2$, define $\delta_i = \delta_i(C, \epsilon, \omega)$ as

$$\delta_0 = \delta - C\epsilon - \omega, \delta_1 = \delta - 2C\epsilon - \omega \text{ and } \delta_2 = \delta - \log(1 - \epsilon) + \omega. \quad (3.1)$$

For any $\omega > 0, \epsilon \in (0, 1)$, we have that $\delta_2(C, \epsilon, \omega) > 0$.

Non giant components estimate

For $i \geq 1$, let \mathcal{E}_i denote the component containing the vertex i . To study non-trivial components of \mathcal{E}_i , we let $T_1 = 1$ and as before, for $r \geq 2$ we let T_r be the number of labelled trees on r vertices. We have the following Lemma that obtains upper and lower bounds concerning the size of non giant components, i.e., components whose size is at most ϵn for some constant $\epsilon > 0$.

Lemma 3.1. *Fix $C \neq 1$ and fix $\epsilon, \omega > 0$ so that both δ_0 and δ_2 defined in (3.1) are positive. There is a positive constant $N = N(C, \epsilon, \omega)$ such that the following two statements hold. For all $n \geq N, 1 \leq i \leq n$ and $1 \leq r \leq \epsilon n$, we have*

$$\mathbb{P}(\#\mathcal{E}_i = r) \leq \frac{T_r e^{-r}}{C(r-1)!} e^{-\delta_0 r}. \quad (3.2)$$

For all $n \geq N, 1 \leq i \leq n$ and $1 \leq r \leq \epsilon n$, we have

$$\mathbb{P}(\#\mathcal{E}_i = r) \geq \frac{1}{C(1-\epsilon)} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_2 r} e^{-2\epsilon + \frac{2\omega}{3}}. \quad (3.3)$$

We need the lower bound in (3.3) to estimate the fraction of nodes present in the giant component for $C > 1$.

Proof of Lemma 3.1: Let $C_u = C + \alpha_n$, $C_d = C - \alpha_n$ and $p_u = \frac{C_u}{n}$ and $p_d = \frac{C_d}{n}$.

We prove the upper bound in (3.2) first. If the vertex i is isolated, then every edge containing i as an endvertex is closed. Since every edge is closed with probability at most $1 - p_d$ and at least $1 - p_u$, we have

$$(1 - p_u)^{n-1} \leq \mathbb{P}(\#\mathcal{E}_i = 1) \leq (1 - p_d)^{n-1}. \quad (3.4)$$

For components with larger size, we argue for $i = 1$ and then generalize for all i . Suppose now that $i = 1$ and the component \mathcal{E}_1 contains $r \geq 2$ vertices. There is a random tree \mathcal{J} contained in \mathcal{E}_1 with the same vertex set as \mathcal{E}_1 and containing $r - 1$ edges, each of which is open. Moreover, every edge with one endvertex belonging to \mathcal{J} and the other endvertex not in \mathcal{J} , is closed. The number of such edges is $r(n - r)$. Let \mathcal{T}_r denote the set of all trees with vertex set $\{1, 2, \dots, r\}$. For a fixed tree $\mathcal{T} \in \mathcal{T}_r$, we therefore have

$$\mathbb{P}(\{\mathcal{E}_1 = \{1, 2, \dots, r\}\} \cap \{\mathcal{T} \subseteq \mathcal{E}_1\}) \leq p_u^{r-1} (1 - p_d)^{r(n-r)}. \quad (3.5)$$

The final expression is because every edge in the graph K_n is open with probability at most p_u and closed with probability at most $1 - p_d$ (see (1.1)). Summing over all possible choices of \mathcal{T} , we have

$$\mathbb{P}(\{\mathcal{E}_1 = \{1, 2, \dots, r\}\}) \leq T_r p_u^{r-1} (1 - p_d)^{r(n-r)} \quad (3.6)$$

where as before T_r denotes the number of labelled trees on r vertices.

The estimate (3.6) is for a particular choice of vertex set for the component \mathcal{E}_1 . The total number of choices for the vertex set of \mathcal{E}_1 is the number of ways of choosing $r - 1$ vertices (apart from the vertex 1) out of the remaining $n - 1$ vertices. We therefore have from (3.6) that

$$\mathbb{P}(\#\mathcal{E}_1 = r) \leq \binom{n-1}{r-1} T_r p_u^{r-1} (1 - p_d)^{r(n-r)}.$$

The above argument holds for all $1 \leq i \leq n$ and so

$$\mathbb{P}(\#\mathcal{E}_i = r) \leq \binom{n-1}{r-1} T_r p_u^{r-1} (1 - p_d)^{r(n-r)}. \quad (3.7)$$

Using $p_u = \frac{C_u}{n}$ we evaluate the product of the first and third terms as

$$\binom{n-1}{r-1} p_u^{r-1} = \frac{(n-1)!}{(r-1)!(n-r)!} \frac{C_u^{r-1}}{n^{r-1}} \leq \frac{C_u^{r-1}}{(r-1)!}. \quad (3.8)$$

The final estimate holds since $\frac{(n-1)!}{(n-r)!n^{r-1}} = \frac{(n-1)\dots(n-r+1)}{n^{r-1}} \leq 1$.

To evaluate the last term in (3.7), we again use $1 - x \leq e^{-x}$ for $x > 0$ to get

$$(1 - p_d)^{r(n-r)} \leq \exp\left(-\frac{C_d r(n-r)}{n}\right) = e^{-C_d r} \exp\left(\frac{C_d r^2}{n}\right). \quad (3.9)$$

Substituting (3.8) and (3.9) into (3.7) we get

$$\mathbb{P}(\#\mathcal{E}_i = r) \leq T_r \frac{C_u^{r-1}}{(r-1)!} e^{-C_d r} \exp\left(\frac{C_d r^2}{n}\right). \quad (3.10)$$

For $r \leq \epsilon n$, we have $\frac{C_d r^2}{n} \leq \epsilon C_d r$ and therefore

$$\mathbb{P}(\#\mathcal{E}_i = r) \leq T_r \frac{C_u^{r-1} e^{-C_d r}}{(r-1)!} e^{C_d \epsilon r} \quad (3.11)$$

Using $C_u = C + \alpha_n$ and $C_d = C - \alpha_n$, we have that

$$\begin{aligned} C_u^{r-1} e^{-C_d r} e^{C_d \epsilon r} &= C^{r-1} (1 + \alpha_n C^{-1})^{r-1} e^{-Cr} e^{\alpha_n r} e^{C\epsilon r} e^{-\epsilon \alpha_n r} \\ &\leq C^{r-1} e^{\alpha_n C^{-1}(r-1)} e^{-Cr} e^{\alpha_n r} e^{C\epsilon r} e^{-\epsilon \alpha_n r} \\ &\leq C^{r-1} e^{-Cr} e^{C\epsilon r} e^{\alpha_n C^{-1}(r-1)} e^{\alpha_n r}. \end{aligned} \quad (3.12)$$

The middle inequality is because $1 + x \leq e^x$ for all $x > 0$.

Since $r \geq 1$, we have that $\alpha_n C^{-1}(r-1) \leq \alpha_n C^{-1}r$ and so we have

$$\alpha_n C^{-1}(r-1) + \alpha_n r \leq (C^{-1} + 1)\alpha_n r \leq \omega r \quad (3.13)$$

for all n large. Here $\omega > 0$ is as in the statement of the Lemma and the final estimate above holds since $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ and so $(C^{-1} + 1)\alpha_n < \omega$ for all n large.

Substituting (3.13) into (3.12), we have that

$$C_u^{r-1} e^{-C_d r} e^{C_d \epsilon r} \leq C^{r-1} e^{-Cr} e^{C\epsilon r} e^{\omega r} = \frac{1}{C} e^{-\delta_0 r} e^{-r} \quad (3.14)$$

for all n large. Here $\delta_0 > 0$ is as in (3.1). Substituting (3.14) into (3.11) we get (3.2).

Proof of (3.3): The proof is analogous as above. As before, we argue for $i = 1$ and then generalize for all i . Let \mathcal{T}_r denote the set of labelled trees with vertex set $\{1, 2, \dots, r\}$. We then have that

$$\mathbb{P}(\mathcal{E}_1 = \{1, 2, \dots, r\}) \geq \sum_{\mathcal{T} \in \mathcal{T}_r} \mathbb{P}(\mathcal{E}_1 = \mathcal{T}). \quad (3.15)$$

For any fixed tree $\mathcal{T} \in \mathcal{T}_r$, suppose $\{\mathcal{E}_1 = \mathcal{T}\}$ occurs. We then have that the $r - 1$ edges of \mathcal{T} are open and every edge with one vertex in \mathcal{T} and one vertex outside is closed. The number of such edges is $r(n - r)$. Moreover, since \mathcal{E}_1 has exactly $r - 1$ edges, the remaining $\binom{r}{2} - r + 1$ edges in K_r are closed. Here K_r is the complete graph with vertex set $\{1, 2, \dots, r\}$. Using the fact that edges are open with probability at least p_d and closed with probability at least $1 - p_u$, we have that

$$\mathbb{P}(\mathcal{E}_1 = \mathcal{T}) \geq p_d^{r-1} (1 - p_u)^{r(n-r) + \binom{r}{2} - r + 1} \quad (3.16)$$

and since the probability is the same for any tree in \mathcal{T}_r , we substitute (3.16) into (3.15) to get

$$\mathbb{P}(\mathcal{E}_1 = \{1, 2, \dots, r\}) \geq T_r p_d^{r-1} (1 - p_u)^{r(n-r) + \binom{r}{2} - r + 1} \quad (3.17)$$

where T_r is the number of labelled trees on r vertices.

Since there are $\binom{n-1}{r-1}$ ways to choose the remaining $r - 1$ vertices for \mathcal{E}_1 , we therefore have

$$\mathbb{P}(\#\mathcal{E}_1 = r) \geq \binom{n-1}{r-1} T_r p_d^{r-1} (1 - p_u)^{r(n-r) + \binom{r}{2} - r + 1}.$$

Again setting $T_1 = 1$ and using (3.4), the above bound also holds for $r = 1$. The above argument holds for all i and so we have

$$\mathbb{P}(\#\mathcal{E}_i = r) \geq \binom{n-1}{r-1} T_r p_d^{r-1} (1 - p_u)^{r(n-r) + \binom{r}{2} - r + 1}. \quad (3.18)$$

It remains to simplify (3.18) to get (3.3). Using $p_d = \frac{C_d}{n}$, we have

$$\binom{n-1}{r-1} p_d^{r-1} = \frac{(n-1)!}{(r-1)!(n-r)!} \frac{C_d^{r-1}}{n^{r-1}} \geq \frac{C_d^{r-1}}{(r-1)!} (1 - \epsilon)^{r-1}. \quad (3.19)$$

The final inequality in (3.19) is true since

$$\frac{(n-1)!}{(n-r)!n^{r-1}} = \frac{(n-1) \dots (n-r+1)}{n^{r-1}} \geq \frac{(n-r)^{r-1}}{n^{r-1}} \geq (1-\epsilon)^{r-1}.$$

In the above, the first inequality follows since $n-k \geq n-r$ for $1 \leq k \leq r$ and the second inequality follows from the fact that $r \leq \epsilon n$.

To estimate the last term in (3.18) we use the fact

$$1-x \geq e^{-x-x^2} \quad (3.20)$$

for $0 < x < \frac{1}{2}$. For completeness we give a small proof of (3.20). For $0 < x < \frac{1}{2}$ we have

$$\log(1-x) = -x - R(x) \quad (3.21)$$

where $0 \leq R(x) = \sum_{k \geq 2} \frac{x^k}{k}$. Expanding the series, we have

$$R(x) = \frac{x^2}{2} \left(1 + x \frac{2}{3} + x^2 \frac{2}{4} + \dots \right) \leq \frac{x^2}{2} (1 + x + x^2 + \dots) = \frac{x^2}{2(1-x)} \leq x^2$$

where the final estimate follows since $1-x \geq \frac{1}{2}$. Substituting into (3.21) gives (3.20).

We fix n_0 large so that

$$p_u = \frac{C_u}{n} \leq \frac{C + \alpha_n}{n} \leq \frac{C+1}{n} \leq \frac{1}{2} \quad (3.22)$$

for all $n \geq n_0$. The second inequality is true since $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Using the bound (3.20) for the last term in (3.18) then gives

$$(1-p_u)^{r(n-r)+\binom{r}{2}-r+1} = \left(1 - \frac{C_u}{n}\right)^{nr - \frac{r^2}{2} - \frac{3r}{2} + 1} \geq e^{-I_1 - I_2} \quad (3.23)$$

where

$$I_1 = \frac{C_u}{n} \left(nr - \frac{r^2}{2} - \frac{3r}{2} + 1 \right) \leq \frac{C_u}{n} (nr + 1) = C_u r + \frac{C_u}{n} \quad (3.24)$$

and

$$I_2 = \frac{C_u^2}{n^2} \left(nr - \frac{r^2}{2} - \frac{3r}{2} + 1 \right) \leq \frac{C_u^2}{n^2} (nr + 1) = \frac{C_u^2 r}{n^2} + \frac{C_u^2}{n^2} \leq \frac{C_u^2 \epsilon}{n} + \frac{C_u^2}{n^2}. \quad (3.25)$$

The final estimate is true since $r \leq \epsilon n$.

Substituting the above two estimates into (3.23) we have

$$\begin{aligned} (1 - p_u)^{r(n-r) + \binom{r}{2} - r + 1} &\geq e^{-C_u r} \exp\left(-\frac{C_u}{2n} - \frac{C_u^2 \epsilon}{n} - \frac{C_u^2}{2n^2}\right) \\ &\geq e^{-C_u r} e^{-2\epsilon} \end{aligned} \quad (3.26)$$

for all $n \geq n_1$. Here n_1 does not depend on the choice of r . The final estimate holds since we have from (3.22) that $\frac{C_u}{2n} + \frac{C_u^2 \epsilon}{n} + \frac{C_u^2}{2n^2}$ converges to zero as $n \rightarrow \infty$.

Using (3.26) and (3.19) in (3.18) gives

$$\mathbb{P}(\#\mathcal{E}_i = r) \geq T_r \frac{C_d^{r-1}}{(r-1)!} (1 - \epsilon)^{r-1} e^{-C_u r} e^{-2\epsilon}. \quad (3.27)$$

Using $C_u = C + \alpha_n$ and $C_d = C - \alpha_n$, we have that

$$\begin{aligned} C_d^{r-1} e^{-C_u r} &= C^{r-1} (1 - \alpha_n C^{-1})^{r-1} e^{-Cr} e^{-\alpha_n r} \\ &\geq C^{r-1} e^{-\alpha_n C^{-1}(r-1) - \alpha_n^2 C^{-2}(r-1)} e^{-Cr} e^{-\alpha_n r} \\ &= C^{r-1} e^{-Cr} e^{-\alpha_n C^{-1}(r-1) - \alpha_n^2 C^{-2}(r-1) - \alpha_n r} \end{aligned} \quad (3.28)$$

for all $n \geq n_2$. Here n_2 does not depend on the choice of r . The middle inequality is obtained using (3.20) since $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$ and so $\alpha_n < \frac{1}{2}$ for all n large.

Fixing $\omega > 0$ as in the statement of the Lemma, we also have that

$$\max(\alpha_n C^{-1}, \alpha_n^2 C^{-2}, \alpha_n) < \frac{\omega}{3}$$

for all $n \geq N_3$. Here N_3 does not depend on the choice of r . Thus the exponent in the final term of (3.28) can be bounded as

$$\alpha_n C^{-1}(r-1) + \alpha_n^2 C^{-2}(r-1) + \alpha_n r \leq \frac{\omega}{3}(r-1) + \frac{\omega}{3}(r-1) + \frac{\omega}{3}r = \omega r - \frac{2\omega}{3}.$$

Substituting the above into (3.28) we have

$$C_d^{r-1} e^{-C_u r} \geq C^{r-1} e^{-Cr} e^{-\omega r} e^{\frac{2\omega}{3}} \quad (3.29)$$

and using this in (3.27) gives (3.3). ■

Cross term estimates

We also need the following estimate on the cross terms to determine the variance in the size of the giant component.

Lemma 3.2. *Fix $C \neq 1$ and fix $\epsilon, \omega > 0$ such that δ_0 and δ_1 defined in (3.1) are positive. There is an integer $N = N(C, \epsilon, \omega)$ so that the following two statements hold. For all $n \geq N, 1 \leq i \neq j \leq n$ and $1 \leq r_1, r_2 \leq \epsilon n$, we have*

$$\begin{aligned} & \mathbb{P}(\{\#\mathcal{E}_i = r_1\} \cap \{\#\mathcal{E}_j = r_2\} \cap \{\mathcal{E}_i \neq \mathcal{E}_j\}) \\ & \leq \frac{1}{C} \frac{T_{r_1} e^{-r_1}}{(r_1 - 1)!} e^{-\delta_1 r_1} \frac{1}{C} \frac{T_{r_2} e^{-r_2}}{(r_2 - 1)!} e^{-\delta_1 r_2}. \end{aligned} \quad (3.30)$$

Similarly, for all $n \geq N, 1 \leq i \neq j \leq n$ and $2 \leq r_1 \leq \epsilon n$, we have

$$\mathbb{P}(\{\#\mathcal{E}_i = r_1\} \cap \{\mathcal{E}_i = \mathcal{E}_j\}) \leq 2\epsilon \frac{1}{C} \frac{T_{r_1} e^{-r_1}}{(r_1 - 1)!} e^{-\delta_0 r_1}. \quad (3.31)$$

Proof of Lemma 3.2: The proof is analogous to the proof of the upper bound (3.2) of Lemma 3.1. As in the proof of Lemma 3.1, let $C_u = C + \alpha_n$, $C_d = C - \alpha_n$ and $p_u = \frac{C_u}{n}$ and $p_d = \frac{C_d}{n}$.

Also we let $i = 1$ and $j = 2$ throughout the proof and the argument holds for all i and j . Let

$$E := \{\#\mathcal{E}_1 = r_1\} \cap \{\#\mathcal{E}_2 = r_2\} \cap \{\mathcal{E}_1 \neq \mathcal{E}_2\} \quad (3.32)$$

and suppose that $\mathcal{E}_1 = \{1, 2, \dots, r_1\}$ and $\mathcal{E}_2 = \{r_1 + 1, \dots, r_1 + r_2\}$ for $r_1, r_2 \geq 2$. Let \mathcal{T}_{r_1} be the set of all trees with vertex set $\{1, 2, \dots, r_1\}$ and let \mathcal{T}_{r_1, r_2} be the set of all trees with vertex set $\{r_1 + 1, \dots, r_1 + r_2\}$. Since \mathcal{E}_1 has $\{1, 2, \dots, r_1\}$ as its vertices, there is a random tree $\mathcal{J}_1 \in \mathcal{T}_{r_1}$ containing $r_1 - 1$ edges, each of which is open. Moreover, every edge with one endvertex in \mathcal{J}_1 and one endvertex outside, is closed. The number of such edges is $r_1(n - r_1)$. Similarly, there is a random tree $\mathcal{J}_2 \in \mathcal{T}_{r_1, r_2}$ containing $r_2 - 1$ edges, each of which is open. Again every edge with one endvertex in \mathcal{J}_2 and one endvertex outside, is closed. In this case, we only need to count the edges not having an endvertex in the tree \mathcal{J}_1 . The number of such edges is $r_2(n - r_1 - r_2)$.

Let

$$E_{r_1, r_2} = E \cap \{\mathcal{E}_1 = \{1, 2, \dots, r_1\}\} \cap \{\mathcal{E}_2 = \{r_1 + 1, \dots, r_1 + r_2\}\}$$

and fix trees $\mathcal{T}_1 \in \mathcal{T}_{r_1}$ and $\mathcal{T}_2 \in \mathcal{T}_{r_1, r_2}$. We have from the discussion in the previous paragraph that

$$\mathbb{P}(E_{r_1, r_2} \cap \{\mathcal{T}_1 \subseteq \mathcal{E}_1\} \cap \{\mathcal{T}_2 \subseteq \mathcal{E}_2\}) \leq A_1 \quad (3.33)$$

where

$$A_1 = A_1(r_1, r_2) = p_u^{r_1-1} (1 - p_d)^{r_1(n-r_1)} p_u^{r_2-1} (1 - p_d)^{r_2(n-r_1-r_2)}. \quad (3.34)$$

In obtaining the above expression, we use the fact that every edge in the complete graph K_n is open with probability at most p_u and closed with probability at most $1 - p_d$ (see (1.1)). Summing over all possible choices of \mathcal{T}_1 and \mathcal{T}_2 we have

$$\mathbb{P}(E_{r_1, r_2}) \leq T_{r_1} T_{r_2} A_1 \quad (3.35)$$

where as before T_r denotes the number of labelled trees on r vertices.

The expression (3.35) obtains estimates for a particular choice of vertex sets for the components \mathcal{E}_1 and \mathcal{E}_2 . To count the total number of choices for the vertex sets of the components \mathcal{E}_1 and \mathcal{E}_2 , we argue as follows. Since $1 \in \mathcal{E}_1$ and $2 \in \mathcal{E}_2$, we choose $r_1 - 1$ vertices for \mathcal{E}_1 and $r_2 - 1$ vertices for \mathcal{E}_2 out of the remaining $n - 2$ vertices. The number of such choices is

$$\binom{n-2}{r_1-1, r_2-1} := \frac{(n-2)!}{(r_1-1)!(r_2-1)!(n-r_1-r_2)!}. \quad (3.36)$$

Using (3.35) and the definition of the event E in (3.32) we therefore have that

$$\mathbb{P}(E) \leq \frac{(n-2)!}{(r_1-1)!(r_2-1)!(n-r_1-r_2)!} T_{r_1} T_{r_2} A_1. \quad (3.37)$$

Setting $T_1 = 1$, we have that the above expression also holds if $r_1 = 1$ or $r_2 = 1$. Substituting the expression for A_1 from (3.34) and rearranging terms, we have

$$\begin{aligned} \mathbb{P}(E) &\leq \frac{(n-2)!}{(r_1-1)!(r_2-1)!(n-r_1-r_2)!} T_{r_1} T_{r_2} p_u^{r_1+r_2-2} (1 - p_d)^{A_2} \\ &= \frac{T_{r_1}}{(r_1-1)!} \frac{T_{r_2}}{(r_2-1)!} \frac{(n-2)!}{(n-r_1-r_2)!} p_u^{r_1+r_2-2} (1 - p_d)^{A_2} \end{aligned} \quad (3.38)$$

where

$$A_2 = A_2(r_1, r_2) = r_1(n - r_1) + r_2(n - r_1 - r_2). \quad (3.39)$$

We evaluate the last two terms in (3.38) separately. Using $p_u = \frac{C_u}{n}$, we have

$$\begin{aligned} \frac{(n-2)!}{(n-r_1-r_2)!} p_u^{r_1+r_2-2} &= (n-2) \dots (n-r_1-r_2+1) \frac{C_u^{r_1+r_2-2}}{n^{r_1+r_2-2}} \\ &= C_u^{r_1+r_2-2} \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{r_1+r_2-1}{n}\right) \\ &\leq C_u^{r_1+r_2-2}. \end{aligned} \quad (3.40)$$

To evaluate the last term in (3.38), we use (3.39) to get

$$(1-p_d)^{A_2} \leq \exp\left(-\frac{C_d A_2}{n}\right) = e^{-C_d(r_1+r_2)} \exp\left(\frac{C_d}{n}(r_1^2 + r_1 r_2 + r_2^2)\right). \quad (3.41)$$

To obtain the first inequality we use the estimate $1-x \leq e^{-x}$ with $x = p_d = \frac{C_d}{n}$. Substituting (3.41) and (3.40) into (3.38) we get

$$\mathbb{P}(E) \leq T_{r_1} \frac{C_u^{r_1-1} e^{-C_d r_1}}{(r_1-1)!} T_{r_2} \frac{C_u^{r_2-1} e^{-C_d r_2}}{(r_2-1)!} e^{A_3}$$

where

$$A_3 = \frac{C_d}{n}(r_1^2 + r_1 r_2 + r_2^2) \leq C_d \epsilon (r_1 + r_1 + r_2) \leq 2C_d \epsilon (r_1 + r_2).$$

The middle estimate is obtained using $r_1 \leq \epsilon n$ and $r_2 \leq \epsilon n$. Thus

$$\mathbb{P}(E) \leq \frac{T_{r_1}}{(r_1-1)!} \frac{T_{r_2}}{(r_2-1)!} A_4(r_1) A_4(r_2) \quad (3.42)$$

where $A_4(r) = C_u^{r-1} e^{-C_d r} e^{2C_d \epsilon r}$.

Arguing as in the derivation of (3.14), we have that

$$A_4(r) = C_u^{r-1} e^{-C_d r} e^{2C_d \epsilon r} \leq C^{r-1} e^{-Cr} e^{2C \epsilon r} e^{\omega r} = \frac{e^{-r}}{C} e^{-\delta_1 r} \quad (3.43)$$

for all $n \geq n_1$, where n_1 does not depend on the choice of r . Here $\omega > 0$ is as in the statement of the Lemma and $\delta_1 > 0$ is as in (3.1). Substituting (3.43) into (3.42) gives (3.30).

The proof of (3.31) is analogous. First we write

$$\mathbb{P}(\{\#\mathcal{E}_1 = r_1\} \cap \{\mathcal{E}_1 = \mathcal{E}_2\}) = \mathbb{P}(\{\#\mathcal{E}_1 = r_1\} \cap \{2 \in \mathcal{E}_1\}).$$

If $\{\#\mathcal{E}_1 = r_1\} \cap \{2 \in \mathcal{E}_1\}$ occurs, then we only need to choose $r_1 - 2$ remaining vertices out of the possible $n - 2$ vertices. The number of such ways is $\binom{n-2}{r_1-2}$. As before, there are T_{r_1} possible labelled trees with r_1 vertices. Arguing as in the proof of (3.2), we have

$$\begin{aligned} & \mathbb{P}(\{\#\mathcal{E}_1 = r_1\} \cap \{2 \in \mathcal{E}_1\}) \\ & \leq \binom{n-2}{r_1-2} T_{r_1} p_u^{r_1-1} (1-p_d)^{r_1(n-r_1)} \\ & = \frac{r_1-1}{n-1} \left(\binom{n-1}{r_1-1} T_{r_1} p_u^{r_1-1} (1-p_d)^{r_1(n-r_1)} \right). \end{aligned} \quad (3.44)$$

The term within the brackets is exactly the term in the right hand side of (3.7) with r replaced by r_1 and whose estimate is obtained as the right hand side of (3.2). Also since $r_1 \leq \epsilon n$, we have that $\frac{r_1-1}{n-1} \leq \frac{\epsilon n-1}{n-1} \leq 2\epsilon$ for all n large.

Combining the above, we have

$$\mathbb{P}(\{\#\mathcal{E}_1 = r_1\} \cap \{2 \in \mathcal{E}_1\}) \leq 2\epsilon \frac{1}{C} \frac{T_{r_1} e^{-r_1}}{(r_1-1)!} e^{-\delta_0 r_1}.$$

This proves (3.31). ■

4 Mid size components estimate

In Lemmas 3.1 and 3.2 of the previous section, we have obtained probability estimates for non giant components; i.e., components of size at most ϵn for some constant $\epsilon > 0$. The next step is to obtain probability estimates for mid size components; i.e., components whose size lies in the range $[M \log n + 1, \epsilon n]$ for some constant $M > 0$. To do so, we use an auxiliary result. Let $T_1 = 1$ and for integer $r \geq 2$, let T_r be the number of labelled trees on r vertices. For $C > 0$, define $q(C)$ as in (1.6) and $\delta = \delta(C) \geq 0$ be as in (1.2). Define

$$R_n(C) := \sum_{r=n+1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta r} = \sum_{r=n+1}^{\infty} \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!}, \quad (4.1)$$

to be the remainder term of $q(C)$. We have

$$\lim_n R_n(C) = 0. \quad (4.2)$$

We directly use the properties of the random graphs to prove the result.

Proof of (4.2): For $1 \leq i \leq n$, we recall that \mathcal{E}_i denotes the component of the random graph G containing the vertex i and so the term

$$Z_n(\epsilon) := \sum_{i=1}^n \mathbf{1}(1 \leq \#\mathcal{E}_i \leq \epsilon n) \quad (4.3)$$

denotes the sum of sizes of all components each of whose size is at most ϵn .

For $C \neq 1$, we fix $\epsilon, \omega > 0$ so that $\delta_2 = \delta_2(C, \epsilon, \omega)$ defined in (3.1) is positive. We then use the lower bound (3.3) in Lemma 3.1 to get

$$\mathbb{E}Z_n(\epsilon) = \sum_{i=1}^n \mathbb{P}(1 \leq \#\mathcal{E}_i \leq \epsilon n) \geq n \sum_{r=1}^{\epsilon n} \frac{1}{C(1-\epsilon)} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_2 r} e^{-2\epsilon + \frac{2\omega}{3}}. \quad (4.4)$$

Fix integer $N \geq 1$. Choose n large so that $\epsilon n \geq N$. We then have from (4.4) that

$$\sum_{r=1}^N \frac{1}{C(1-\epsilon)} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_2 r} e^{-2\epsilon + \frac{2\omega}{3}} \leq \frac{1}{n} \mathbb{E}Z_n(\epsilon) \leq 1. \quad (4.5)$$

The final estimate is true since $Z_n(\epsilon) \leq n$. Also $\delta_2(C, \epsilon, \epsilon) = \delta - (2C + 1)\epsilon$ is a decreasing function of ϵ and is positive for all $\epsilon > 0$ small. Therefore allowing $\epsilon = \omega \downarrow 0$ in (4.5) gives

$$\sum_{r=1}^N \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta r} \leq 1 \quad (4.6)$$

where $\delta > 0$ is as in (1.2). Allowing $N \rightarrow \infty$ gives $q(C) \leq 1$ for all $C \neq 1$.

For $C = 1$, we again use (4.6). Allowing $C \uparrow 1$ in (4.6) gives

$$\sum_{r=1}^N \frac{T_r e^{-r}}{(r-1)!} \leq 1. \quad (4.7)$$

Again allowing $N \rightarrow \infty$ gives $q(1) \leq 1$. Thus

$$q(C) \leq 1 \quad (4.8)$$

for all $C > 0$ and this proves (4.2). ■

The following Lemma establishes that with high probability there are no components with size greater than a constant multiple of $\log n$ and less than ϵn .

For any $M > 0$ and $\epsilon > 0$, let

$$B(M, \epsilon) := \bigcup_{1 \leq i \leq n} \{M \log n + 1 \leq \#\mathcal{E}_i \leq \epsilon n\} \quad (4.9)$$

denote the event that there is a component whose size lies in the range $[M \log n + 1, \epsilon n]$. Let $\delta = \delta(C)$ be as in (1.2). We have the following Lemma.

Lemma 4.1. *Fix $C \neq 1$ and $\epsilon > 0$ so that δ_0 defined in (3.1) is positive. Fix $M > \frac{1}{\delta_0}$ and $\theta < M\delta_0 - 1$. There is a positive constant $N = N(C, \epsilon, M, \theta)$ such that*

$$\mathbb{P}(B(M, \epsilon)) \leq \frac{1}{n^\theta} \quad (4.10)$$

for all $n \geq N$.

Proof of Lemma 4.1: For $M \log n + 1 \leq r \leq \epsilon n$ we have that

$$e^{-\delta_0 r} \leq \exp(-M\delta_0 \log n) = \frac{1}{n^{M\delta_0}}.$$

Therefore from (3.2) we have for any $1 \leq i \leq n$ that

$$\mathbb{P}(\#\mathcal{E}_i = r) \leq \frac{T_r e^{-r}}{C(r-1)!} e^{-\delta_0 r} \leq \frac{T_r e^{-r}}{C(r-1)!} \frac{1}{n^{M\delta_0}}.$$

Hence

$$\begin{aligned} \mathbb{P}(M \log n + 1 \leq \#\mathcal{E}_i \leq \epsilon n) &\leq \frac{1}{n^{M\delta_0}} \sum_{r=M \log n + 1}^{\epsilon n} \frac{T_r e^{-r}}{C(r-1)!} \\ &\leq \frac{1}{C n^{M\delta_0}} R_{M \log n}(1) \end{aligned}$$

where $R_{M \log n}(\cdot)$ is the remainder term defined in (4.1). Using (4.2) we have that $R_{M \log n}(1) \leq C$ for all n large. Thus we have

$$\mathbb{P}(M \log n + 1 \leq \#\mathcal{E}_i \leq \epsilon n) \leq \frac{1}{n^{M\delta_0}}$$

for all $n \geq n_0$. Here n_0 does not depend on the choice of i .

Using the above bound, we have

$$\mathbb{P}(B(M, \epsilon)) \leq \sum_{i=1}^n \mathbb{P}(M \log n + 1 \leq \#\mathcal{E}_i \leq \epsilon n) \leq \frac{n}{n^{M\delta_0}}.$$

If $\theta < M\delta_0 - 1$, then the right hand size is at most $\frac{1}{n^\theta}$ for all n large. \blacksquare

5 Mean and Variance estimates for $Z_n(\epsilon)$

In this section, we obtain mean and variance estimates for the term $Z_n(\epsilon)$ defined in (4.3), representing the number of vertices present in non giant components, i.e. components whose size is at most ϵn .

We first have some preliminary estimates. Fix $\epsilon, \omega > 0$ such that $\delta_i = \delta_i(C, \epsilon, \omega)$ defined in (3.1) is positive. By definition $\delta_i(C, \eta_1, \eta_2)$ is positive for all $0 < \eta_1 < \epsilon$ and $0 < \eta_2 < \omega$. For $i = 0, 1, 2$, let

$$q_i(C, \epsilon, \omega) := e^{-C} + \sum_{r=2}^{\infty} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_i r} \quad (5.1)$$

We have the following result. Fix $C \neq 1$ and $i \in \{0, 1, 2\}$. We have

$$\lim_{\epsilon \downarrow 0} q_i(C, \epsilon, \epsilon) = q(C). \quad (5.2)$$

Proof of (5.2): We first prove for $i = 2$. Let $\epsilon_m \downarrow 0$ be any sequence. Since $(1 - \epsilon)e^\epsilon \leq 1$ for any $\epsilon > 0$, we have

$$q_2(C, \epsilon_m, \epsilon_m) = \sum_{r=1}^{\infty} b_r (1 - \epsilon_m)^r e^{\epsilon_m r} \leq \sum_{r=1}^{\infty} b_r = q(C)$$

where $b_r = \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!}$ for $r \geq 1$. Therefore $\limsup_m q_2(C, \epsilon_m, \epsilon_m) \leq q(C)$. We obtain the lower bound as follows. For any fixed integer $N \geq 1$ we have

$$q_2(C, \epsilon_m, \epsilon_m) = \sum_{r=1}^{\infty} b_r (1 - \epsilon_m)^r e^{\epsilon_m r} \geq \sum_{r=1}^N b_r (1 - \epsilon_m)^r e^{\epsilon_m r}$$

Allowing $m \rightarrow \infty$ and then $N \rightarrow \infty$ gives

$$\liminf_m q_2(C, \epsilon_m, \epsilon_m) \geq q(C). \quad (5.3)$$

Therefore $\lim_m q_2(C, \epsilon_m, \epsilon_m) = q(C)$ and (5.2) holds for $i = 2$.

We now see that (5.2) holds for $i = 1$ and an analogous proof holds for $i = 0$. We again let $\epsilon_m \downarrow 0$ be any sequence and obtain

$$q_1(C, \epsilon_m, \epsilon_m) = \sum_{r=1}^{\infty} b_r e^{\beta_m r} \geq \sum_{r=1}^{\infty} b_r = q(C)$$

where $\beta_m = 2C\epsilon_m + \epsilon_m$. Therefore $\liminf_m q_1(C, \epsilon_m, \epsilon_m) \geq q(C)$. To obtain the upper bound, we use the fact that $\delta_1(C, \epsilon_m, \epsilon_m) = \delta - \beta_m > 0$ for all $m \geq M_0$ large. Fix $m \geq M_0$ and any integer $N \geq 1$. We then have that

$$q_1(C, \epsilon_m, \epsilon_m) = \sum_{r=1}^{\infty} b_r e^{\beta_m r} = \sum_{r=1}^N b_r e^{\beta_m r} + R_{1,N}(C, \epsilon_m) \quad (5.4)$$

where

$$R_{1,N}(C, \epsilon_m) := \sum_{r=N+1}^{\infty} b_r e^{\beta_m r} = \sum_{r=N+1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} e^{-(\delta-\beta_m)r} \leq \sum_{r=N+1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} \quad (5.5)$$

and the last term is precisely $R_N(1)$ defined in (4.1).

Substituting the above into (5.4) we have

$$q_1(C, \epsilon_m, \epsilon_m) \leq \sum_{r=1}^N b_r e^{\beta_m r} + R_N(1)$$

for all $m \geq M_0$ and for any fixed integer $N \geq 1$. Allowing $m \rightarrow \infty$ in the above and using $\beta_m \rightarrow 0$, we have

$$\limsup_m q_1(C, \epsilon_m, \epsilon_m) \leq \sum_{r=1}^N b_r + R_N(1).$$

Allowing $N \rightarrow \infty$ in the above and using (4.2), we get

$$\limsup_m q_1(C, \epsilon_m, \epsilon_m) \leq \sum_{r \geq 1} b_r = q(C).$$

Thus $\lim_m q_1(C, \epsilon_m, \epsilon_m) = q(C)$ and this proves (5.2) for $i = 1$. ■

Mean estimates for $Z_n(\epsilon)$

We have the following bounds on the mean of

$$Z_n(\epsilon) = \sum_{1 \leq i \leq n} \mathbf{1}(1 \leq \mathcal{E}_i \leq \epsilon n). \quad (5.6)$$

For $C > 0$, let $q = q(C)$ be as defined in (1.6).

Lemma 5.1. *Fix $0 < \gamma < 1$ and $C \neq 1$. There is a positive constant $\epsilon_1 = \epsilon_1(C, \gamma)$ so that the following statement holds for all $0 < \epsilon < \epsilon_1$. There is a positive constant $N = N(C, \epsilon, \gamma)$ so that for all $n \geq N$, we have*

$$nq(C)(1 - \gamma) \leq \mathbb{E}Z_n(\epsilon) \leq nq(C)(1 + \gamma). \quad (5.7)$$

Proof of Lemma 5.1: Fix $i \in \{0, 2\}$ and let $q_i(\cdot, \cdot, \cdot)$ be as defined in (5.1). From (5.2), we have that $q_i(C, \eta, \eta) \rightarrow q(C)$ as $\eta \rightarrow 0$. For any fixed $\eta > 0$, the term $\delta_2 = \delta_2(C, \eta, \eta) = \delta - \log(1 - \eta) + \eta$ is positive. Here $\delta = \delta(C) > 0$ is as in (1.2). Also, the term $\delta_0 = \delta_0(C, \eta, \eta) = \delta - (2C + 1)\eta > 0$ for all $0 < \eta < \frac{\delta}{2C+1}$.

For a fixed $0 < \gamma < 1$, we therefore let $\epsilon_1 = \epsilon_1(C, \gamma) > 0$ be small so that for all $0 < \epsilon < \epsilon_1$, the following statements hold. The term

$$\left(1 - \frac{3\gamma}{4}\right) \frac{e^{-\frac{4\epsilon}{3}}}{(1 - \epsilon)} \geq 1 - \gamma. \quad (5.8)$$

The terms $\delta_1(C, \epsilon, \epsilon)$ and $\delta_2(C, \epsilon, \epsilon)$ are positive and

$$q_0(C, \epsilon, \epsilon) \leq q(C)(1 + \gamma) \text{ and } q_2(C, \epsilon, \epsilon) \geq q(C) \left(1 - \frac{\gamma}{2}\right). \quad (5.9)$$

Fix an $\epsilon > 0$ so that the above statements hold. We obtain an upper bound for $Z_n(\epsilon)$ first. Using the upper bound (3.2) in Lemma 3.1, we have that

$$\frac{1}{n} \mathbb{E}Z_n(\epsilon) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(1 \leq \# \mathcal{E}_i \leq \epsilon n) \leq \sum_{r=1}^{\epsilon n} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_0 r} \leq q_0(C, \epsilon, \epsilon). \quad (5.10)$$

Using (5.9), we have

$$\frac{1}{n} \mathbb{E}Z_n(\epsilon) \leq q(C)(1 + \gamma). \quad (5.11)$$

To obtain a lower bound on $\mathbb{E}Z_n(\epsilon)$, we use the lower bound (3.3) in Lemma 3.1 with $\epsilon = \omega$ to get

$$\begin{aligned}
\frac{1}{n}\mathbb{E}Z_n(\epsilon) &= \frac{1}{n} \sum_{i=1}^n \mathbb{P}(1 \leq \#\mathcal{E}_i \leq \epsilon n) \\
&\geq \sum_{r=1}^{\epsilon n} \frac{1}{C(1-\epsilon)} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_2 r} e^{-\frac{4\epsilon}{3}} \\
&= (q_2(C, \epsilon, \epsilon) - R_{2,n}(C, \epsilon)) \frac{e^{-\frac{4\epsilon}{3}}}{(1-\epsilon)} \tag{5.12}
\end{aligned}$$

where

$$0 \leq R_{2,n}(C, \epsilon) = \sum_{r=\epsilon n+1}^{\infty} \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!} e^{-\delta_2 r} \leq \sum_{r=\epsilon n+1}^{\infty} \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!} = R_{\epsilon n}(C).$$

Here $R_n(\cdot)$ is the remainder term as in (4.1).

Since the remainder $R_{\epsilon n}(C) \rightarrow 0$ as $n \rightarrow \infty$ (see (4.2)), we have

$$R_{\epsilon n}(C) \leq q(C) \frac{\gamma}{4}$$

for all large n . Substituting the above and the estimate for $q_2(C, \epsilon, \epsilon)$ (see (5.9)) into (5.12) we have

$$\frac{1}{n}\mathbb{E}Z_n(\epsilon) \geq q(C) \left(1 - \frac{3\gamma}{4}\right) \frac{e^{-\frac{4\epsilon}{3}}}{(1-\epsilon)} \geq q(C)(1-\gamma) \tag{5.13}$$

by our choice of $\epsilon > 0$ in (5.8). Combining (5.11) and (5.13) gives (5.7). ■

Variance estimate for $Z_n(\epsilon)$

We have the following estimate on the variance of $Z_n(\epsilon)$.

Lemma 5.2. *Fix $0 < \gamma < 1$ and $C \neq 1$. There is a positive constant $\epsilon_2 = \epsilon_2(C, \gamma)$ so that the following statement holds for all $0 < \epsilon < \epsilon_2$. There is a positive integer $N = N(C, \epsilon, \gamma)$ so that for all $n \geq N$, we have*

$$\text{var}(Z_n(\epsilon)) \leq nq(C)(1+\gamma) + 4n^2q^2(C)\gamma$$

where $q(C)$ is as in (1.6).

Proof of Lemma 5.2: For $i = 0, 1, 2$ let $\delta_i = \delta_i(C, \epsilon, \epsilon)$. For $1 \leq i \leq n$, let $X_i = \mathbf{1}(1 \leq \#\mathcal{E}_i \leq \epsilon n)$ denote the indicator function of the event that the component containing vertex i has size at most ϵn . From (4.3) we have that $Z_n(\epsilon) = \sum_{i=1}^n X_i$ and so

$$\mathbb{E}Z_n^2(\epsilon) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}X_i X_j = \sum_{i=1}^n \mathbb{E}X_i^2 + \sum_{i \neq j} \mathbb{E}(X_i X_j). \quad (5.14)$$

Since each X_i is an indicator function, we have that $X_i^2 = X_i$ and so the first term in (5.14) is $\sum_{i=1}^n \mathbb{E}X_i = \mathbb{E}Z_n(\epsilon)$. We evaluate the second term in (5.14) using the cross term estimates in Lemma 3.2. For $1 \leq i \neq j \leq n$ we write

$$E(X_i X_j) = \mathbb{E}(X_i X_j \mathbf{1}(\mathcal{E}_i = \mathcal{E}_j)) + \mathbb{E}(X_i X_j \mathbf{1}(\mathcal{E}_i \neq \mathcal{E}_j)) \quad (5.15)$$

and evaluate each term separately.

Using the cross term estimate (3.30) of Lemma 3.2 we have

$$\begin{aligned} \mathbb{E}(X_i X_j \mathbf{1}(\mathcal{E}_i \neq \mathcal{E}_j)) &= \sum_{r_1=1}^{\epsilon n} \sum_{r_2=1}^{\epsilon n} \mathbb{P}(\{\#\mathcal{E}_i = r_1\} \cap \{\#\mathcal{E}_j = r_2\} \cap \{\mathcal{E}_i \neq \mathcal{E}_j\}) \\ &\leq \sum_{r_1=1}^{\epsilon n} \sum_{r_2=1}^{\epsilon n} \frac{1}{C} \frac{T_{r_1} e^{-r_1}}{(r_1 - 1)!} e^{-\delta_1 r_1} \frac{1}{C} \frac{T_{r_2} e^{-r_2}}{(r_2 - 1)!} e^{-\delta_1 r_2} \\ &\leq q_1^2(C, \epsilon, \epsilon) \end{aligned} \quad (5.16)$$

where $q_1(\cdot, \cdot, \cdot)$ is as in (5.1).

Using (3.31) of Lemma 3.2, we similarly have

$$\begin{aligned} \mathbb{E}(X_i X_j \mathbf{1}(\mathcal{E}_i = \mathcal{E}_j)) &= \sum_{r_1=2}^{\epsilon n} \mathbb{P}(\{\#\mathcal{E}_i = r_1\} \cap \{\mathcal{E}_i = \mathcal{E}_j\}) \\ &\leq \sum_{r_1=2}^{\epsilon n} 2\epsilon \frac{1}{C} \frac{T_{r_1} e^{-r_1}}{(r_1 - 1)!} e^{-\delta_0 r_1} \\ &\leq 2\epsilon q_0(C, \epsilon, \epsilon) \end{aligned} \quad (5.17)$$

Combining (5.16) and (5.17) and substituting in (5.15) we have

$$E(X_i X_j) \leq q_1^2(C, \epsilon, \epsilon) + 2\epsilon q_0(C, \epsilon, \epsilon). \quad (5.18)$$

The right hand side of (5.18) converges to $q^2(C)$ as $\epsilon \rightarrow 0$ using (5.2). Here $q(C)$ is as defined in (1.6). Fix $\epsilon_2 = \epsilon_2(C, \gamma) > 0$ small so that $\epsilon_2 < \epsilon_1(C, \gamma)$

and the right hand side of (5.18) is at most $q^2(C)(1 + \gamma)^2$. Here $\epsilon_1(.,.)$ is as defined in the statement of Lemma 5.1.

Fixing $0 < \epsilon < \epsilon_2$ we therefore have $\sum_{i \neq j} \mathbb{E}X_i X_j \leq n(n-1)q^2(C)(1 + \gamma)$. This evaluates the second term in (5.14). To evaluate the first term in (5.14), we use the fact that each $X_i, 1 \leq i \leq n$ is an indicator function. Therefore $X_i^2 = X_i$ and the first term in (5.14) is $\sum_{i=1}^n \mathbb{E}X_i = \mathbb{E}Z_n(\epsilon)$. Using the above estimates for the two terms of (5.14) we have

$$\mathbb{E}Z_n^2(\epsilon) \leq \mathbb{E}Z_n(\epsilon) + n(n-1)q^2(C)(1 + \gamma)^2 \leq nq(C)(1 + \gamma) + n^2q^2(C)(1 + \gamma)^2$$

where the final estimate is obtained using the upper bound in (5.7) of Lemma 5.1. This is possible by our choice of $\epsilon > 0$. Using the lower bound in (5.7), we therefore have

$$\begin{aligned} \text{var}(Z_n(\epsilon)) &= \mathbb{E}Z_n^2(\epsilon) - (\mathbb{E}Z_n(\epsilon))^2 \\ &\leq nq(C)(1 + \gamma) + n^2q^2(C)(1 + \gamma)^2 - n^2q^2(C)(1 - \gamma)^2 \\ &= nq(C)(1 + \gamma) + 4n^2q^2(C)\gamma. \end{aligned}$$

This proves the lemma. ■

6 Subcritical case

For the subcritical case, there are two proofs. If we assume property (ii) of Theorem 3 that $q(C) = 1$ for $0 < C < 1$, we then obtain a weaker version of Theorem 1 using Chebychev's inequality and the mean and variance estimates of Lemmas 5.1 and 5.2. In Section 8, we provide such a proof for the supercritical case $C > 1$.

In what follows we prove Theorem 1 *without assuming* Theorem 3. We use the result of Theorem 1 to give a probabilistic proof of properties (i) – (iii) of Theorem 3 in the next Section.

Fixing $C < 1$, we have from Lemma 4.1 that with high probability there are no mid size components whose size lies in the range $[M \log n + 1, \epsilon n]$. To obtain decay for giant components, i.e., components of size larger than ϵn , we proceed as follows: We first show that with high probability, all vertices in the component \mathcal{E}_i containing the vertex $i, 1 \leq i \leq n$, are within a distance of order of $\log n$ from the vertex i and then see that it is not possible to contain order of n vertices within such short distance.

For integer $t \geq 1$, let $\mathcal{N}_t(i)$ denote the set of vertices at a distance of t from vertex i in the random graph G . Therefore $\#\mathcal{N}_t(i)$ denotes the number of vertices at a distance t from vertex i . Here distance between vertices x and y refers to the graph distance and is the number of edges in the path with the least number of edges between x and y . Since there are at most n vertices in \mathcal{E}_i , we have that $1 \leq t \leq n$. Recalling the definition of the sequence α_n from (1.1), we have the following Lemma.

Lemma 6.1. *Fix $C < 1$ and let $C_u = C + \alpha_n$. There is an integer $N \geq 1$ so that for all $n \geq N$ we have that $C_u < 1$. For all $n \geq N, 1 \leq i \leq n$ and $1 \leq t \leq n$, we have*

$$\mathbb{E}(\#\mathcal{N}_t(i)) \leq C_u^t \quad (6.1)$$

and

$$\mathbb{E}((\#\mathcal{N}_t(i))^2) \leq \frac{C_u^t}{1 - C_u}. \quad (6.2)$$

Proof of Lemma 6.1: We first prove (6.1). We set $i = 1$ throughout and for $1 \leq t \leq n$ define $\mathcal{N}_t := \mathcal{N}_t(1)$. The proof holds for arbitrary $1 \leq i \leq n$. Let $S_0 = \{1\}$ and $U_0 = \{1, 2, \dots, n\} \setminus S_0$. Fixing $t \geq 1$ we have

$$\mathbb{E}(\#\mathcal{N}_t) = \sum_{S_0=\{1\}, S_1, \dots, S_{t-1}} \mathbb{E}(\#\mathcal{N}_t \mathbf{1}(\mathcal{N}_{t-1} = S_{t-1}, \dots, \mathcal{N}_1 = S_1, \mathcal{N}_0 = S_0)) \quad (6.3)$$

where the summation is over all subsets S_1, \dots, S_{t-1} , of $\{1, 2, \dots, n\}$. For fixed S_1, \dots, S_{t-1} , define the event

$$F_{t-1} := \{\mathcal{N}_{t-1} = S_{t-1}, \dots, \mathcal{N}_1 = S_1, \mathcal{N}_0 = S_0\} \quad (6.4)$$

and the set

$$U_{t-1} = \{1, 2, \dots, n\} \setminus (\cup_{i=0}^{t-1} S_i). \quad (6.5)$$

If the sets $\{S_i\}$ are not mutually disjoint, we have that $\mathbf{1}(F_{t-1}) = 0$. If however the indicator function $\mathbf{1}(F_{t-1}) = 1$, then the vertices in \mathcal{N}_t satisfy the following properties:

- (i) If $v \in \mathcal{N}_t$, then v is adjacent (i.e. connected by an open edge) to some vertex in S_{t-1} . Also, the vertex v is not adjacent to any vertex in S_i for $0 \leq i \leq t-1$.
- (ii) All vertices of \mathcal{N}_t are in U_{t-1} .

Proof of (i) – (ii): The first statement of (i) is true as follows. Fix $v \in \mathcal{N}_t$. There is a path $(e(1, i_1), e(i_1, i_2), \dots, e(i_{t-1}, i_t = v))$ consisting of t open edges

from the vertex 1 to vertex v . Here we represent the open edge joining the vertices i and j as $e(i, j)$. The vertex i_{t-1} is at a distance of $t-1$ from vertex 1 and therefore $i_{t-1} \in S_{t-1}$. For the second statement, we argue as follows. If the vertex v is adjacent to some vertex in S_i for some $1 \leq i \leq t-2$, then the distance between the vertex 1 and the vertex v is at most $t-1$. This is a contradiction since the vertex v is at a distance of t from vertex 1.

For proving (ii), is true, we suppose that $v \notin U_{t-1}$. From the definition (6.5) for U_{t-1} , we then have that $v \in \cup_{i=0}^{t-1} S_i$, a contradiction to property (i) proved above. ■

From properties (i) – (ii) above we have that

$$\#\mathcal{N}_t \mathbf{1}(F_{t-1}) \leq \sum_{y \in S_{t-1}} \sum_{z \in U_{t-1}} X_{y,z} \mathbf{1}(F_{t-1}) = \sum_{y \in S_{t-1}} J_y \mathbf{1}(F_{t-1}) \quad (6.6)$$

where $X_{y,z}$ denotes the indicator function of the event that the edge between vertices y and z is open and $J_y = \sum_{z \in U_{t-1}} X_{y,z}$ for $y \in S_{t-1}$. We have an upper bound in (6.6) since a single vertex $z \in U_{t-1}$ can be connected to multiple vertices in S_{t-1} . Also we have that the following property.

$$\begin{aligned} &\text{The event } F_{t-1} \text{ defined in (6.4) is independent of} \\ &\text{the indicator functions } \{X_{y,z}\}_{y \in S_{t-1}, z \in U_{t-1}}. \end{aligned} \quad (6.7)$$

Proof of (6.7): For subsets $A, B \subseteq \{1, 2, \dots, n\}$, let (A, B) denote the set of all edges with one endvertex in A and other endvertex in B . For integer $i \geq 0$, define the event

$$V_i = \bigcap_{e \in (S_i, S_{i+1})} \{e \text{ is open}\} \bigcap \bigcap_{e \in (S_i, U_i \setminus S_{i+1})} \{e \text{ is closed}\}.$$

The event V_i depends only on the set of edges having an endvertex in the set S_i .

The event $F_1 = \{N_1 = S_1, N_0 = S_0\}$ can be written as $F_1 = V_0$ and so the event F_1 depends only on the state of edges containing $S_0 = \{1\}$ as the endvertex. Similarly, the event

$$F_2 = \{N_2 = S_2\} \cap F_1 = V_1 \cap V_0$$

depends only on the state of the edges that have an endvertex in $S_0 \cup S_1$. In particular, the event F_2 does not depend on the state of edges having both

endvertices in $U_1 = \{1, 2, \dots, n\} \setminus (S_0 \cup S_1)$. Since $S_2 \subset U_1$ and $U_2 \subset U_1$, any edge with one endvertex in S_2 and other endvertex in U_2 has both endvertices in U_1 . Therefore the event F_2 does not depend on the state of edges having one endvertex in S_2 and other endvertex in U_2 . Continuing this way inductively, the event $F_{t-1} = \{N_{t-1} = S_{t-1}\} \cap F_{t-2}$ does not depend on the state of edges having one endvertex in S_{t-1} and other endvertex in U_{t-1} . ■

Setting $p_u = \frac{C_u}{n}$ we have from (6.6) and (6.7) that

$$\mathbb{E} \# \mathcal{N}_t \mathbf{1}(F_{t-1}) \leq \sum_{y \in S_{t-1}} \mathbb{E} J_y \mathbf{1}(F_{t-1}) = \sum_{y \in S_{t-1}} \mathbb{E} J_y \mathbb{P}(F_{t-1}) \leq C_u \# S_{t-1} \mathbb{P}(F_{t-1}) \quad (6.8)$$

where $J_y = \sum_{z \in U_{t-1}} X_{y,z}$ is as defined in (6.6). The final inequality is obtained because, for any fixed $y \in S_{t-1}$ we have that

$$\mathbb{E}(J_y) = \sum_{z \in U_{t-1}} \mathbb{E} X_{y,z} \leq \# U_{t-1} p_u \leq n p_u = C_u. \quad (6.9)$$

Substituting (6.8) into (6.3) we have

$$\begin{aligned} \mathbb{E}(\# \mathcal{N}_t) &\leq C_u \sum_{S_0=\{1\}, S_1, \dots, S_{t-1}} \# S_{t-1} \mathbb{P}(F_{t-1}) \\ &= C_u \sum_{S_0=\{1\}, S_1, \dots, S_{t-1}} \# S_{t-1} \mathbb{P}(\mathcal{N}_{t-1} = S_{t-1}, \dots, \mathcal{N}_1 = S_1, \mathcal{N}_0 = S_0) \\ &= C_u \mathbb{E}(\# \mathcal{N}_{t-1}). \end{aligned} \quad (6.10)$$

Continuing this iteratively we get $\mathbb{E}(\# \mathcal{N}_t) \leq C_u^t \mathbb{E}(\# \mathcal{N}_0) = C_u^t$. This proves (6.1).

The proof of (6.2) is analogous. Indeed proceeding as before, we have

$$\mathbb{E}(\# \mathcal{N}_t)^2 = \sum_{S_0=\{1\}, S_1, \dots, S_{t-1}} \mathbb{E}((\# \mathcal{N}_t)^2 \mathbf{1}(\mathcal{N}_{t-1} = S_{t-1}, \dots, \mathcal{N}_1 = S_1, \mathcal{N}_0 = S_0)). \quad (6.11)$$

Defining F_{t-1} as in (6.4) and using (6.7), we have

$$\mathbb{E}(\# \mathcal{N}_t)^2 \mathbf{1}(F_{t-1}) \leq \mathbb{E} \left(\sum_{y \in S_{t-1}} J_y \right)^2 \mathbf{1}(F_{t-1}) = \mathbb{E} \left(\sum_{y \in S_{t-1}} J_y \right)^2 \mathbb{P}(F_{t-1}) \quad (6.12)$$

where $J_y = \sum_{z \in U_{t-1}} X_{y,z}$ is as defined in (6.6).

If $y_1 \neq y_2$, the random variables J_{y_1} and J_{y_2} are independent and so

$$\begin{aligned} \mathbb{E} \left(\sum_{y \in S_{t-1}} J_y \right)^2 &= \text{var} \left(\sum_{y \in S_{t-1}} J_y \right) + \left(\mathbb{E} \sum_{y \in S_{t-1}} J_y \right)^2 \\ &= \sum_{y \in S_{t-1}} \text{var}(J_y) + \left(\sum_{y \in S_{t-1}} \mathbb{E} J_y \right)^2. \end{aligned} \quad (6.13)$$

The mean $\mathbb{E} J_y \leq C_u$ using (6.9) and regarding the variance, we have that

$$\text{var}(J_y) = \sum_{z \in U_{t-1}} \text{var}(X_{y,z}) \leq \sum_{z \in U_{t-1}} \mathbb{E} X_{y,z}^2 = \sum_{z \in U_{t-1}} \mathbb{E} X_{y,z} \leq C_u. \quad (6.14)$$

The first equality in (6.14) holds since the random variables $\{X_{y,z}\}_{z \in U_{t-1}}$ are independent. The second equality in (6.14) holds since the random variable $X_{y,z}$ takes the value either 0 or 1. The final estimate in (6.14) follows from (6.9).

Substituting the estimates (6.9) and (6.14) into (6.13) we get

$$\mathbb{E} \left(\sum_{y \in S_{t-1}} J_y \right)^2 \leq C_u \#S_{t-1} + (C_u \#S_{t-1})^2.$$

and using the above in (6.12) we have

$$\mathbb{E}(\#\mathcal{N}_t)^2 \mathbf{1}(F_{t-1}) \leq (C_u \#S_{t-1} + C_u^2 (\#S_{t-1})^2) \mathbb{P}(F_{t-1}).$$

From the expression (6.11), we then have

$$\begin{aligned} \mathbb{E}(\#\mathcal{N}_t)^2 &\leq \sum_{S_0=\{1\}, S_1, \dots, S_{t-1}} (C_u \#S_{t-1} + C_u^2 (\#S_{t-1})^2) \mathbb{P}(F_{t-1}) \\ &= C_u \mathbb{E}(\#\mathcal{N}_{t-1}) + C_u^2 \mathbb{E}(\#\mathcal{N}_{t-1})^2 \\ &\leq C_u^t + C_u^2 \mathbb{E}(\#\mathcal{N}_{t-1})^2 \end{aligned} \quad (6.15)$$

where the final estimate follows from (6.1).

Setting $a_t = \mathbb{E}(\#\mathcal{N}_t)^2$ and proceeding iteratively using (6.15), we have

$$a_t \leq \sum_{j=t}^{2t-1} C_u^j + C_u^{2t} a_0 = \sum_{j=t}^{2t} C_u^j$$

since $a_0 = \mathbb{E}\#(\mathcal{N}_0)^2 = 1$. Therefore we have that

$$a_t \leq \sum_{j \geq t} C_u^j = \frac{C_u^t}{1 - C_u}.$$

This proves (6.2). ■

Large components estimates

Using Lemma 6.1, we see that large components cannot exist with high probability.

Lemma 6.2. *Fix $C < 1$. For every $\epsilon > 0$, there are positive constants $D = D(C, \epsilon)$ and $N = N(C, \epsilon)$ such that*

$$\mathbb{P}(\#\mathcal{E}_i \geq \epsilon n) \leq D \frac{(\log n)^3}{n^2} \quad (6.16)$$

for all $n \geq N$ and all $1 \leq i \leq n$.

Proof of Lemma 6.2: As in the proof of Lemma 6.1, we set $i = 1$ throughout and let \mathcal{N}_t denote the (random) set of vertices at distance $t \geq 1$ from vertex 1. The proof holds for any $1 \leq i \leq t$. Let

$$\tau = \inf\{t \geq 1 : \mathcal{N}_t = \emptyset\} \quad (6.17)$$

denote the largest distance of a vertex from the vertex 1. Fixing $C < C_1 < 1$, we have that $C_u = C + \alpha_n \leq C_1$ for all $n \geq n_0(C, C_1)$. The above statement is true since $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$. Using (6.1), we have for $t = \frac{-2}{\log C_1} \log n$ that

$$\mathbb{P}(\tau \geq t) \leq \mathbb{P}(\#\mathcal{N}_t \geq 1) \leq \mathbb{E}\#\mathcal{N}_t \leq C_u^t = \exp\left(-2 \frac{\log C_u}{\log C_1} \log n\right) \leq \frac{1}{n^2} \quad (6.18)$$

where the final inequality follows since $C_u \leq C_1 < 1$ and so $\frac{\log C_u}{\log C_1} \geq 1$. Thus for $\epsilon > 0$, we have

$$\begin{aligned} \mathbb{P}(\#\mathcal{E}_1 \geq \epsilon n) &= \mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau \geq t\}) + \mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau < t\}) \\ &\leq \mathbb{P}(\{\tau \geq t\}) + \mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau < t\}) \\ &\leq \frac{1}{n^2} + \mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau < t\}) \end{aligned} \quad (6.19)$$

using (6.18).

To evaluate the second term in (6.19), we write

$$\mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau < t\}) = \sum_{k=1}^{t-1} \mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau = k\}) \quad (6.20)$$

and fix $1 \leq k \leq t-1$. If the event $\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau = k\}$ occurs, then some $\mathcal{N}_j, 1 \leq j \leq k$ contains at least $\frac{\epsilon n}{k} \geq \frac{\epsilon n}{t}$ vertices and so

$$\mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau = k\}) \leq \mathbb{P}\left(\bigcup_{j=1}^k \{\#\mathcal{N}_j \geq \epsilon n t^{-1}\}\right) \leq \sum_{j=1}^k \mathbb{P}(\#\mathcal{N}_j \geq \epsilon n t^{-1}). \quad (6.21)$$

For any fixed $1 \leq j \leq k$, we have that

$$\mathbb{P}(\#\mathcal{N}_j \geq \epsilon n t^{-1}) \leq \frac{t^2}{(\epsilon n)^2} \mathbb{E}(\#\mathcal{N}_j)^2 \leq \frac{t^2}{(\epsilon n)^2} \frac{C_u^j}{1 - C_u} \quad (6.22)$$

where the first inequality follows using Markov inequality and the second inequality follows from the estimate (6.2).

Using (6.22), the final term in (6.21) can be bounded above by

$$\sum_{j=1}^k \frac{t^2}{(\epsilon n)^2} \frac{C_u^j}{1 - C_u} \leq \sum_{j \geq 1} \frac{t^2}{(\epsilon n)^2} \frac{C_u^j}{1 - C_u} = \frac{t^2}{\epsilon^2 n^2} \frac{C_u}{(1 - C_u)^2} \leq D_1 \frac{t^2}{n^2}, \quad (6.23)$$

for all $n \geq n_1$ and some constant $D_1 = D_1(C, \epsilon) > 0$. Here n_1 does not depend on the choice of k . The final estimate holds since $C_u = C + \alpha_n \rightarrow C < 1$ as $n \rightarrow \infty$ and so $\frac{C_u}{(1 - C_u)^2}$ is a bounded sequence in n . The final term in (6.23) is an estimate for $\mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau = k\})$ and holds for any $1 \leq k \leq t-1$. Substituting the above estimate into (6.20) we have

$$\mathbb{P}(\{\#\mathcal{E}_1 \geq \epsilon n\} \cap \{\tau < t\}) \leq t D_1 \frac{t^2}{n^2} \leq D_2 \frac{(\log n)^3}{n^2} \quad (6.24)$$

for some constant $D_2 = D_2(C, C_1, \epsilon) > 0$ and all $n \geq n_2$ large. The final estimate holds since $t = -\frac{2}{\log C_1} \log n$. Finally, substituting (6.24) into (6.19) we get (6.16). \blacksquare

Proof of Theorem 1: Fix $\epsilon > 0$ small so that $\delta_0(C, \epsilon, \epsilon)$ defined in (3.1) is positive and let $\delta_0 = \delta_0(C, \epsilon, \epsilon)$.

Let

$$W(\epsilon) := \bigcup_{1 \leq i \leq n} \{\#\mathcal{E}_i \geq \epsilon n\} \quad (6.25)$$

denote the event that there is a component of size larger than ϵn . From (6.16) of Lemma 6.2, we have that

$$\mathbb{P}(W(\epsilon)) \leq \sum_{i=1}^n \mathbb{P}(\#\mathcal{E}_i \geq \epsilon n) \leq D \frac{(\log n)^3}{n} \quad (6.26)$$

where the constant D is as in (6.16). This estimates that large components cannot exist with high probability.

To see existence of mid size components, we use Lemma 4.1. Let $\delta > 0$ be as in (1.2) and let $M > \frac{1}{\delta}$ be fixed and fix $\theta < \min(M\delta - 1, 1)$. Since $\delta_0 = \delta - C\epsilon - \epsilon \rightarrow \delta$ as $\epsilon \rightarrow 0$, we fix $\epsilon > 0$ small enough so that $M > \frac{1}{\delta_0}$ and $\theta < \min(M\delta_0 - 1, 1)$. Fix $\theta_1 > 0$ so that $\theta < \theta_1 < \min(M\delta_0 - 1, 1)$ and let $B(M, \epsilon)$ be the event defined in (4.9) that there is a component whose size lies in the range $[M \log n + 1, \epsilon n]$. From (4.10) we have

$$\mathbb{P}(B(M, \epsilon)) \leq \frac{1}{n^{\theta_1}} \quad (6.27)$$

for all n large.

If $B^c(M, \epsilon) \cap W^c(\epsilon)$ occurs, then every component has size at most $M \log n$. In other words, the event $H_1(M)$ defined in (1.3) occurs. From (6.27) and (6.26), we have that

$$\mathbb{P}(H_1) \geq 1 - \frac{1}{n^{\theta_1}} - D_2 \frac{(\log n)^3}{n} \geq 1 - \frac{1}{n^\theta}$$

for all n large. This proves (1). ■

7 Proof of Theorem 3

To prove Theorem 3, we need the following properties of the function $q(C)$ defined in (1.6).

- (a1) The function $q(C)$ is continuous at any $C > 0$.
- (a2) If $C_1 > C_2 \geq 1$, then $q(C_1) < q(C_2)$ strictly.

Proof of (a1) – (a2): To prove the continuity property (a1), we use an approximation procedure and truncate the series expansion for $q(C)$ at a finite number of terms. For $C > 0$, we let

$$q_{0,N}(C) = \sum_{r=1}^N \frac{T_r C^{r-1} e^{-Cr}}{(r-1)!}$$

and have $q_{0,N}(C) \uparrow q(C)$ as $N \rightarrow \infty$. The function $q_{0,N}(C)$ is continuous function of C for any fixed N . Let $C_0 > 0$ be fixed. We have

$$\begin{aligned} |q(C_0) - q(C)| &\leq |q(C_0) - q_{0,N}(C_0)| + |q_{0,N}(C_0) - q_{0,N}(C)| \\ &\quad + |q_{0,N}(C) - q(C)| \\ &= (q(C_0) - q_{0,N}(C_0)) + |q_{0,N}(C_0) - q_{0,N}(C)| \\ &\quad + (q_{0,N}(C) - q(C)) \\ &= R_N(C_0) + |q_{0,N}(C_0) - q_{0,N}(C)| + R_N(C) \end{aligned}$$

where the second equality follows from the fact that $q_{0,N}(\cdot) \leq q(\cdot)$ by definition. The term $R_N(\cdot)$ is the remainder term defined in (4.1). We also note that for any $C > 0$, we have

$$R_N(C) = \sum_{r=N+1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta r} \leq \sum_{r=N+1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} = R_N(1).$$

where $\delta = \delta(C) \geq 0$ is as in (1.2). Thus

$$|q(C_0) - q(C)| \leq |q_{0,N}(C_0) - q_{0,N}(C)| + 2R_N(1) \quad (7.1)$$

Fix $C_0 > 0$ and let $\{C_m\}_{m \geq 1}$ be any sequence such that $C_m > 0$ for all $m \geq 1$ and $C_m \rightarrow C_0$ as $m \rightarrow \infty$. Fix integer $N \geq 1$. Using (7.1) and the continuity of $q_{0,N}$ we have

$$\limsup_m |q(C_0) - q(C_m)| \leq 2R_N(1).$$

Allowing $N \rightarrow \infty$ and using (4.2) then gives $\lim_m q(C_m) = q(C_0)$. This proves that $q(C)$ is a continuous function for $C > 0$.

To see the strictly decreasing property (a2), we fix $C_1 > C_2 \geq 1$. We then have $C_1 e^{-C_1} < C_2 e^{-C_2}$ and $\frac{1}{C_1} < \frac{1}{C_2}$. Therefore for any integer $r \geq 1$, we have

$$C_1^{r-1} e^{-C_1 r} = \frac{1}{C_1} (C_1 e^{-C_1})^r < \frac{1}{C_2} (C_2 e^{-C_2})^r = C_2^{r-1} e^{-C_2 r}.$$

Multiplying by $\frac{T_r}{(r-1)!}$ both sides and summing over r gives that $q(C_1) < q(C_2)$ strictly. This proves the strictly decreasing property of $q(C)$ for $C > 1$. ■

Proof of Theorem 3: (i) Proved before, see (4.8).

(ii) This is true by the property of the extinction probability for the Poisson branching process (see Durrett (2007)). For completeness, however, we give a small proof directly using the properties of random graphs.

Fix $0 < \gamma < 1$ and $C \neq 1$. Let $\epsilon > 0$ be such that (5.7) holds and let $Z_n(\epsilon)$ be as defined in (5.6). We have from (5.7) that

$$\mathbb{E}Z_n(\epsilon) \leq nq(C)(1 + \gamma)$$

for all n large. We also have that $Z_n(\epsilon) = n$ if and only if there is no component of size larger than ϵn i.e., if and only if the event $W^c(\epsilon)$ holds. Here $W(\epsilon)$ is as defined in (6.25). Using the corresponding estimate (6.26) we have that

$$nq(C)(1 + \gamma) \geq \mathbb{E}Z_n(\epsilon) \geq n\mathbb{P}(W^c(\epsilon)) \geq n \left(1 - D_2 \frac{(\log n)^3}{n} \right)$$

where the constant $D_2 > 0$ is as in (6.26). Therefore

$$q(C) \geq \frac{1}{1 + \gamma} \left(1 - D_2 \frac{(\log n)^3}{n} \right)$$

for all n large. Allowing $n \rightarrow \infty$ and then $\gamma \downarrow 0$ gives $q(C) \geq 1$. We have so far proved that $q(C) = 1$ for $0 < C < 1$.

To prove (iii) and that $q(1) = 1$, we use the properties (a1) – (a2) of $q(\cdot)$ described above.

(iii) Since $q(C)$ is continuous at $C = 1$ (see property (a1)), we have

$$\lim_{C \uparrow 1} q(C) = q(1) = 1.$$

Also we have that $q(C)$ is strictly decreasing in C for $C > 1$ (see property (a2)) and so

$$0 < q(C) < q(1) = 1 \tag{7.2}$$

for all $C > 1$. This proves (iii).

(iv) The proof is combinatorial and we refer to Chapter 14, van Lint and Wilson (2010). ■

8 Supercritical case

Proof of Theorem 4: We recall that $Z_n(\epsilon)$ defined in (5.6) denotes the sum of sizes of components each of whose size is at most ϵn . We use the estimates on mean and variance of $Z_n(\epsilon)$ in Lemma 5.1 and Lemma 5.2 to see that at least one giant component exists.

Fix $0 < \gamma < 1, C > 1$ and let $\delta > 0$ be as in (1.2). Define $\epsilon_3 = \min(\epsilon_1, \epsilon_2)$ where ϵ_1 and ϵ_2 are defined in Lemma 5.1 and 5.2, respectively. Let $q = q(C)$ be as in (1.6).

Using the Chebychev's inequality we have

$$\mathbb{P}\left(|Z_n(\epsilon) - \mathbb{E}Z_n(\epsilon)| > q\frac{\gamma}{2}n\right) \leq \frac{4\text{var}(Z_n(\epsilon))}{q^2\gamma^2n^2} \quad (8.1)$$

where $\text{var}(Z_n(\epsilon)) = \mathbb{E}(Z_n(\epsilon) - \mathbb{E}Z_n(\epsilon))^2$ is the variance of the random variable $Z_n(\epsilon)$. Since $0 < \gamma < 1$ is arbitrary, we use Lemma 5.2 with $\gamma_1 = \frac{\gamma^3}{64}$ to get that $\text{var}(Z_n(\epsilon)) \leq nq\left(1 + \frac{\gamma^3}{64}\right) + n^2q^2\frac{\gamma^3}{16}$ for all n large. Therefore the final term in (8.1) is bounded above by

$$\frac{4}{q^2\gamma^2n^2} \left(nq\left(1 + \frac{\gamma^3}{64}\right) + n^2q^2\frac{\gamma^3}{16} \right) = \frac{4}{q\gamma^2n} \left(1 + \frac{\gamma^3}{64} \right) + \frac{\gamma}{4} \leq \frac{\gamma}{2} \quad (8.2)$$

for all n large.

Thus if $Y_n(\epsilon) = n - Z_n(\epsilon)$ denotes the sum of sizes of components each of which has size at least ϵn as defined in (1.12), we have from (8.1) and (8.2) that

$$\mathbb{P}\left(|Y_n(\epsilon) - \mathbb{E}Y_n(\epsilon)| > q\frac{\gamma}{2}n\right) \leq \frac{\gamma}{2} \quad (8.3)$$

for all n large. Since $q = q(C) < q(1) = 1$ strictly for $C > 1$ (see properties (ii) – (iii) of Theorem 3), we have that there exists at least one giant component. Using Lemma 5.1 with $\frac{\gamma}{2}$, we also have

$$n\left(1 - q - \frac{\gamma}{2}\right) \leq \mathbb{E}Y_n(\epsilon) \leq n\left(1 - q + \frac{\gamma}{2}\right) \quad (8.4)$$

for all n large. If the estimate (8.4) holds and the event on the left hand side of (8.3) holds, then the event $V(\gamma, \epsilon)$ defined in (1.13) occurs. We therefore have from (8.3) and (8.4) that

$$\mathbb{P}(V(\gamma, \epsilon)) \geq 1 - \frac{\gamma}{2} \quad (8.5)$$

for all n large.

Next, we recall the event $B(M, \epsilon)$ defined in (4.9) that there exists a component in the range $[M \log n + 1, \epsilon n]$. Let $\delta > 0$ be as in (1.2). Fix $M > \frac{1}{\delta}$ and choose ϵ_3 (defined in the first paragraph of this proof) smaller if necessary so that the following holds for all $0 < \epsilon < \epsilon_3$. If $\delta_0 = \delta_0(C, \epsilon, \epsilon) = \delta - (2C + 1)\epsilon$, then $\delta_0 > 0$ and $M > \frac{1}{\delta_0}$. Fixing $0 < \epsilon < \epsilon_3$ and $0 < \theta < \min(M\delta_0 - 1, 1)$, we have from (4.10) that

$$\mathbb{P}(B(M, \epsilon)) \leq \frac{1}{n^\theta} \quad (8.6)$$

for all n large.

From (8.5) and (8.6), we note that

$$\mathbb{P}(V^c(\gamma, \epsilon) \cup B(M, \epsilon)) \leq \frac{\gamma}{2} + \frac{1}{n^\theta} \leq \gamma \quad (8.7)$$

for all n large.

If $V(\gamma, \epsilon) \cap B^c(M, \epsilon)$ occurs, then there is at least one component containing at least ϵn nodes. Also, since $B^c(M, \epsilon)$ occurs, size of all small components is at most $M \log n$. And by definition, the event $H_2 = H_2(M, \gamma, \epsilon)$ defined in (1.14) occurs. From (8.7), we obtain Theorem 4. \blacksquare

9 Proof of Theorems 5 and 2

Proof of Theorem 5: We first see that there is a unique solution to $\delta_0(C, \frac{1}{2}, 0) = \frac{C}{2} - 1 - \log C = 0$ for $C > 1$. The derivative $\delta'_0 = \delta'_0(C, \frac{1}{2}, 0)$ with respect to C , satisfies $\delta'_0 = \frac{1}{2} - \frac{1}{C} > 0$ for $C > 2$ and so the function $\delta_0(C, \frac{1}{2}, 0)$ is increasing for all $C > 2$. Moreover, $\delta_0(2, \frac{1}{2}, 0) < 0$ and $\delta_0(C, \frac{1}{2}, 0) \rightarrow \infty$ as $C \rightarrow \infty$. Therefore there is a unique $C_0 > 2$ such that $\delta_0(C_0, \frac{1}{2}, 0) = 0$.

To see that there is no solution for $\delta_0(C, \frac{1}{2}, 0) = 0$ in the interval $(1, 2)$, we use the fact that the derivative $\delta'_0 = \frac{1}{2} - \frac{1}{C} < 0$ for all $C < 2$. Thus $\delta_0(C, \frac{1}{2}, 0)$ is decreasing in the interval $(1, 2)$ and $\delta_0(1, \frac{1}{2}, 0) = -\frac{1}{2} < 0$. This implies that there is no solution for $\delta_0(C, \frac{1}{2}, 0) = 0$ in the interval $(1, 2)$. We also have $C_0 < 8$, since $\delta_0(8, \frac{1}{2}, 0) = 3 - \log 8 = 3(1 - \log 2) > 0$.

To obtain the uniqueness of the giant component, we use the upper bound in the tree counting estimate (3.2) of Lemma 3.1. To ensure that the upper bound is valid, we fix $C > C_0$ and let $\epsilon_0 > \frac{1}{2}$ and $\alpha, \omega_0 > 0$ be such that

$$0 < \delta_0(C, \alpha + \epsilon_0, \omega_0) < \delta_0(C, \epsilon_0, \omega_0) \quad (9.1)$$

If $Z_n(\epsilon_0)$ as defined in (5.6) is the number of vertices belonging to non giant components i.e., components with size at most $\epsilon_0 n$, we have from (3.2) that

$$\frac{1}{n} \mathbb{E} Z_n(\epsilon_0) = \frac{1}{n} \sum_{i=1}^n \mathbb{P}(1 \leq \#\mathcal{E}_i \leq \epsilon_0 n) \leq \sum_{r=1}^{\epsilon_0 n} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_0 r} \leq q_0(C, \epsilon_0, \omega_0). \quad (9.2)$$

Here $\delta_0 = \delta_0(C, \epsilon_0, \omega_0)$ and $q_0 := q_0(C, \epsilon_0, \omega_0)$ is as defined in (5.1). Using Markov inequality, we have that

$$\mathbb{P}(Z_n(\epsilon_0) > n e^{\alpha C} q_0) \leq \frac{\mathbb{E} Z_n(\epsilon_0)}{e^{\alpha C} q_0} \leq e^{-\alpha C} \quad (9.3)$$

where $\alpha > 0$ is as in (9.1).

We have that $e^{\alpha C} q_0 < 1$ strictly. To see this is true, we first have from the choice of $\alpha > 0$ in (9.1) that $\delta_\alpha := \delta_0(C, \epsilon_0 + \alpha, \omega_0) > 0$. Also from definition of $q(C)$ in (1.6) and from Theorem 3 we have that $q(1) = \sum_{r=1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} = 1$. Therefore

$$q_0(C, \epsilon_0 + \alpha, \omega_0) = \sum_{r=1}^{\infty} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_\alpha r} \leq \frac{1}{C} \sum_{r=1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} = \frac{1}{C} < 1$$

since $C > 1$. Since $e^{-\delta_\alpha r} = e^{-\delta_0 r} e^{C\alpha r} \geq e^{-\delta_0 r} e^{C\alpha}$ for integer $r \geq 1$, we also have

$$q_0(C, \epsilon_0 + \alpha, \omega) = \sum_{r=1}^{\infty} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_\alpha r} \geq e^{\alpha C} \sum_{r=1}^{\infty} \frac{1}{C} \frac{T_r e^{-r}}{(r-1)!} e^{-\delta_0 r} = e^{\alpha C} q_0.$$

Thus $e^{\alpha C} q_0 < 1$, strictly.

The term $Z_n(\epsilon_0)$ is the sum of sizes of non giant components i.e., the components whose size lies between 1 and $\epsilon_0 n$. If $\{Z_n(\epsilon_0) \leq n e^{\alpha C} q_0\}$ occurs, then $Z_n(\epsilon_0) < n$ strictly and therefore there exists a component with at least $\epsilon_0 n + 1 > \frac{n}{2} + 1$ vertices. This proves the first step in the proof.

The next step is to eliminate mid size components. Since the term $\delta_0 = \delta_0(C, \epsilon_0, \omega_0) > 0$, the result of Lemma 4.1 holds. In particular, for a fixed integer $M_0 > \frac{1}{\delta_0}$ we have from estimate (4.10) of Lemma 4.1 that

$$\mathbb{P}(B(M_0, \epsilon_0)) \leq \frac{1}{n^\theta} \quad (9.4)$$

for some constant $\theta > 0$. Here $B(M_0, \epsilon_0)$ denotes the event that there exists a component whose size lies in the range $[M_0 \log n + 1, \epsilon_0 n]$.

From (9.3) and (9.4), we obtain that

$$\mathbb{P}(\{Z_n(\epsilon_0) \leq ne^{\alpha C} q_0\} \cap B^c(M_0, \epsilon_0)) \geq 1 - e^{-\alpha C} - \frac{1}{n^\theta} \quad (9.5)$$

for all $n \geq n_0$. Here $n_0 \geq 1$ is a constant. If the event on the left hand side occurs, there is at least one component containing at least $\epsilon_0 n > \frac{n}{2} + 10$ vertices, for all n large. This means that every other component has at most $\frac{n}{2} - 10$ vertices. But since $B^c(M_0, \epsilon_0)$ also occurs, every other component has size at most $M_0 \log n$. Therefore there is exactly one (giant) component containing at least $\frac{n}{2} + 10$ vertices and every other component has at most $M_0 \log n$ vertices.

The final step is to use the results of Theorem 4 to obtain a better estimate for the size of the unique giant component. For $0 < \gamma < 1, \epsilon > 0$, let $H_2 = H_2(M, \gamma, \epsilon)$ be as defined in the statement of Theorem 4. Fixing $M > \frac{1}{\delta}$ and choosing $\epsilon > 0$ small we have from (1.4) of Theorem 4 that $\mathbb{P}(H_2) \geq 1 - \frac{\gamma}{2}$. Using estimate (9.5), we therefore have that

$$\mathbb{P}(\{Z_n(\epsilon_0) \leq ne^{\alpha C} q_0\} \cap B^c(M_0, \epsilon_0) \cap H_2) \geq 1 - e^{-\alpha C} - \frac{1}{n^\theta} - \frac{\gamma}{2} \geq 1 - e^{-\alpha C} - \gamma \quad (9.6)$$

for all n large.

Suppose that the event on the left hand side of (9.6) occurs. By the discussion following (9.5), there is exactly one giant component containing at least $\frac{n}{2} + 1$ nodes and the rest of the components have size at most $M_0 \log n$. But since H_2 also occurs, the size of the giant component is in the range $[(1 - q(C) - \gamma)n, (1 - q(C) + \gamma)n]$ and every other component has size at most $M \log n$. Here $q(C)$ is as defined in (1.6). This proves Theorem 5. ■

Proof of Theorem 2: If $C < e^{-1}$, then the term

$$\delta_0(C, \epsilon, 0) = C(1 - \epsilon) - 1 - \log C = C(1 - \epsilon) + \log\left(\frac{e^{-1}}{C}\right) > 0$$

for all $0 < \epsilon \leq 1$. Fixing $\epsilon = 1$, we have that $\delta_0 = \delta_0(C, 1, 0) = \log\left(\frac{e^{-1}}{C}\right) > 0$. Therefore the upper bound (3.2) of Lemma 3.1 holds for all $1 \leq r \leq n$. Fixing $M > 0$, we therefore have that

$$\mathbb{P}(\#\mathcal{E}_i \geq M \log n) \leq \sum_{r=M \log n}^n \frac{T_r e^{-r}}{C(r-1)!} e^{-\delta_0 r} \leq e^{-\delta_0 M \log n} \sum_{r=M \log n}^n \frac{T_r e^{-r}}{C(r-1)!}. \quad (9.7)$$

The summation in the final term is bounded above as

$$\sum_{r=M \log n}^n \frac{T_r e^{-r}}{C(r-1)!} \leq \frac{1}{C} \sum_{r=1}^{\infty} \frac{T_r e^{-r}}{(r-1)!} = \frac{q(1)}{C} = \frac{1}{C}$$

using Theorem 3. Substituting in (9.7) we have

$$\mathbb{P}(\#\mathcal{E}_i \geq M \log n) \leq \frac{1}{C} e^{-\delta_0 M \log n} = \frac{1}{C} \frac{1}{n^{M\delta_0}} \leq \frac{1}{n^{\theta+1}}$$

for all n large, provided $M > 0$ is large. Thus

$$\mathbb{P}\left(\bigcup_{1 \leq i \leq n} \#\mathcal{E}_i \geq M \log n\right) \leq \frac{1}{n^{\theta}}.$$

This implies that with probability at least $1 - \frac{1}{n^{\theta}}$ all the components have size at most $M \log n$. ■

Acknowledgements

I thank Professors Rahul Roy, Thomas Mountford, Siva Athreya and Federico Camia for crucial comments and for my fellowships. I also thank ISF-UGC for my fellowship.

References

- [1] N. Alon and J. Spencer. (2008). *The Probabilistic Method*. *Wiley*.
- [2] B. Bollobas. (2001). *Random Graphs*. *Cambridge Univ. Press*.
- [3] R. Durrett. (2007). *Random Graph Dynamics*. *Cambridge Univ. Press*.
- [4] P. Erdős and A. Rényi. (1960). On the evolution of random graphs. *Publ. Math. Inst., Hungarian Academy of Sciences*.
- [5] G. Grimmett. (1980). Random labelled trees and their branching networks. *J. Austral. Math. Soc. (Series A)*, **30**, 229–237 .

- [6] R. van der Hofstad. (2016). Random Graphs and Complex Networks - Volumes I and II. *Available at <http://www.win.tue.nl/~rhofstad/>*
- [7] S. Janson, T. Luczak and A. Rucinski. (2000). Random Graphs. *Wiley*.
- [8] J. van Lint and R. Wilson. (2010). Combinatorics. *Cambridge Univ. Press*.